WILLIAM E. MAGNUSSON
Coordenação de Pesquisas em Ecologia
Instituto Nacional de Pesquisas da Amazônia

GUILHERME MOURÃO
Embrapa Pantanal
Empresa Brasileira de Pesquisa Agropecuária

ESTATÍSTICA SEM MATEMÁTICA

A ligação entre as questões e as análises

editora
PLANTA

DEDICATÓRIA

Para Graeme Caughley, cuja filosofia permeia este livro, mas que infelizmente não viveu para o enriquecer com seu estilo indomável de escrever.

AGRADECIMENTOS

Muitas pessoas contribuíram para o desenvolvimento deste livro, principalmente nossos alunos. Entretanto, somente podemos mencionar umas poucas pessoas que contribuíram no estágio final da obra. Sua estrutura geral foi concebida durante a estadia de um de nós (W.E.M.) na Universidade de Griffith, Austrália, graças à intervenção de Carla Catterall e Marc Hero e a bolsa de pós-doutorado recebida da CAPES. Mike Dale revisou a versão em Inglês e corrigiu muitos dos deslizes gramaticais, estatísticos e filosóficos. A versão em Português se beneficiou das revisões cuidadosas de Helena Bergallo, Ísis Medri, Agostinho Catella e de uma multidão de estudantes de nossos cursos. O livro teria levado muito mais tempo a ser publicado se não fosse pelo empenho, apoio e confiança do nosso editor Efraim Rodrigues.

ÍNDICE

Capítulo 1	**Introdução**	1
	O que é delineamento amostral?	6
	O que esperamos que você obtenha deste livro	10
Capítulo 2	**Fluxogramas e questões científicas**	15
	Construindo a hipótese inicial	16
	Três tipos de estudo	20
	1. Isto é real?	20
	2. Qual é a forma e magnitude do efeito?	21
	3. O que acontecerá se as condições mudarem?	22
	Qual o tamanho do problema?	23
	Para onde ir?	24
Capítulo 3	**Descrevendo a natureza:convenções "científicas" e algumas técnicas úteis**	25
	Repetições geradas no computador	32
Capítulo 4	**Quanta evidência é necessária?**	33
	Qual a qualidade da informação?	34
Capítulo 5	**Quando *improvável* significa *bem* possível**	41
	Como os livros de texto contam a estória?	47
	Como os estatísticos contam observações independentes?	49
	É mais fácil compreender o mundo se não virá-lo de cabeça para baixo	50
Capítulo 6	**Como evitar acumular o risco em comparações simples**	53
	Com que tipo de risco estamos preocupados?	53
	Usando a variabilidade para reconhecer uma diferença	55
	Uma premissa importante	58
	Partição das variâncias	60
Capítulo 7	**Análises para um mundo com todas as tonalidades**	63
	Encaixotando o mundo	64
	Descrevendo um mundo retilíneo	67
	O quanto o modelo se ajusta?	69

Capítulo 8 **Problemas do mundo real: mais de um fator**
Fatores simultâneos
Adicionando variabilidades repartidas
Checando premissas com gráficos de parciais
Interações

Capítulo 9 **Quais variáveis analisar?**
Inteligência artificial
Variáveis fantasmas geradas por computadores

Capítulo 10 **Modelos complexos**
Estimando efeitos diretos
Estimando efeitos indiretos
Algumas pedras na análise de caminhos

Capítulo 11 **Endireitando o mundo com transformações e outros truques**
Estimativas por tentativa e erro sem transformação
Outros métodos atípicos
Modelos gerais lineares
Verossimilhança máxima
Problemas e armadilhas da estimativa não-linear

Capítulo 12 **Análise multivariada: cortando as árvores para enxergar melhor a floresta**
Gráficos de gradientes
Gradientes hipotéticos
Mais do que uma dimensão
Análises de vetores "eigen"
A força da cultura: testes de significância
Discriminando entre grupos
Categorias que se ramificam como árvores
Selecionando variáveis
Saber o que queremos antes de começar

Capítulo 13 **Como escrever melhor de trás para frente**
Esquema conceitual simples
Detalhes aborrecidos, mas importantes
Questões do tipo "Por quê"
Uma palavra final

Índice Remissivo de Assuntos
Referências Bibliográficas

CAPÍTULO 1

[A última coisa que o mundo precisa é de mais um livro de estatística]

INTRODUÇÃO

A última coisa que o mundo precisa é de mais um livro de estatística. Existem dezenas deles, a que os estatísticos podem recorrer, quando necessário. Muitos são escritos com estilo e leveza. Então, para quê dois ecólogos, que se sentem especialmente incompetentes em matemática, se arriscariam a escrever um livro que trata de conceitos de estatística? Uma das razões é que temos, já por alguns anos, lecionado um curso um pouco "diferente" de estatística básica, especialmente endereçado para estudantes de pós-graduação em ecologia, e, de alguma forma, este curso tem revolucionado a habilidade destes estudantes em comunicar seus resultados de pesquisa (Magnusson 1997). Entretanto, não usávamos nenhum livro texto para acompanhar este curso e estudantes e professores sempre nos cobravam um. A outra razão é que nos demos conta de que nosso curso vem servindo principalmente para remediar falhas acumuladas na formação dos estudantes (Magnusson 1997). Gastamos um tempo enorme para ensinar conceitos básicos que os estudantes desaprenderam durante seus cursos básicos de estatística.

Tukey (1980) já percebera que "Os estudantes que nunca foram expostos à estatística confirmatória parecem aprender a estatística exploratória mais prontamente." Os maiores erros no delineamento amostral resultam de não se levar em conta conceitos básicos de lógica que muitos estudantes levariam, se sua atenção não tivesse sido desviada pela matemática das estatísticas. Platt (1964) colocou isto de forma eloqüente na seguinte passagem, traduzida um pouco livremente: ["Você pode capturar um fenômeno em uma malha lógica ou matemática. A lógica é uma malha grossa, mas forte. A matemática é fina, porém frágil. A matemática é uma forma bonita de embrulhar um problema, mas não pode reter sua essência, a não ser que ela tenha sido capturada na malha lógica desde o começo".] Guttman (1985) se referiu a isso como "contingente e conteúdo".

É claro que esperamos poder ensinar algum conteúdo através da matemática, porque alguns conceitos estatísticos/matemáticos podem nos ajudar a ver o mundo mais claramente. Contudo, estes não são os conceitos enfatizados nos cursos regulares de estatística. Os estudantes freqüentemente nos perguntam porque os cursos regulares de estatística não tratam destes assuntos. A resposta é que eles tratam. Se você pegar as primeiras páginas de cada capítulo de qualquer bom livro de estatística e colocá-las junto, elas contariam uma história muito semelhante à que contaremos neste livro. Outros autores têm se dado conta da necessidade de passar aos leitores uma visão geral que coloque as diferentes análises estatísticas em uma mesma ordenação lógica. No final do primeiro capítulo de seu livro Harris (1975) escreveu o seguinte: "Para quê ler o resto deste livro? Podemos considerar que se um estudante de doutorado em psicologia entendeu plenamente os conceitos contidos nesta seção, os quais se baseiam apenas em bom senso, então este estudante atingiu cerca de 90% da habilidade necessária para interpretar estatísticas multivariadas." Contudo, poucas pessoas lêem o primeiro capítulo de Harris ou de qualquer outro livro de estatística. Um pesquisador está interessado nas interações de muitos fatores e alguém diz "você precisa de regressão múltipla (ou análise de componentes principais, ou análise de variância fatorial, ou outro procedimento aparentemente complicado), então vá para a página 365." O autor do livro deve ter tido muita dor de cabeça para apresentar a seqüência lógica, que gradativamente levaria ao entendimento necessário para se usar a informação apresentada na página 365. Contudo, poucos irão ler a obra página por página. Nenhum dos autores deste livro leu qualquer outro livro de estatística do princípio ao fim, na ordem em que os autores apresentaram o conteúdo. Mas gostaríamos que vocês lessem este livro desta forma e portanto fizemos nossos capítulos bastante curtos.

Este livro trata da estatística básica e delineamento experimental que os estudantes precisam para entender a literatura ecológica. Quando dizemos "estatística básica" não queremos dizer ficar tirando ao acaso bolinhas coloridas de um saco, ou usar análise de variância para comparar as taxas de crescimento de sorgo em canteiros com três níveis diferentes de fertilizantes. Estas questões não são básicas, são triviais. Em nosso curso usualmente gastamos 3 dias (24 horas-aula) para preparar os alunos para simples comparações de médias, mas isto não é o ponto final. Se depois de 10 dias, o estudante não entender as bases de regressões múltiplas, análise de variância fatorial, estatística multivariada e "análise de caminhos", ele não será capaz de ler a literatura. É preciso estar "alfabetizado" para aprender o conhecimento acadêmico.

Há muitos níveis pelos quais se pode abordar uma matéria e a escolha da abordagem é uma decisão pessoal e criticamente importante. Para ilustrar a *nossa* escolha, vamos apresentar duas analogias bem emocionais. Se você desejasse aprender a respeito de armas de fogo, poderia começar tomando aulas de balística e engenharia de materiais. Ou você poderia ler os folhetos dos fabricantes de armas e aprender como a posse de armas fará você se destacar socialmente e torná-lo(a) mais atraente para o sexo oposto. Entretanto, *nós* começaríamos ensinado que armas de fogo foram projetadas para matar pessoas ou animais e que há considerações éticas e práticas em relação ao seu uso. Um fisiologista pode praticar sexo para obter dados ou amostras, e investigar a química da reprodução. Revistas populares ensinam que o intercurso sexual é um meio de atingir um *status* social elevado e que dispor de muitos parceiros é motivo para orgulho. *Nós* começaríamos

dizendo que o intercurso sexual tem as funções básicas de servir de meio de comunicação entre duas pessoas e, eventualmente, fazer bebês. Não estamos dizendo que nosso ponto de vista é melhor ou mais abrangente que os demais, apenas que ele é o mais importante para nós. Acreditamos que tratados matemáticos sobre estatística são tão importantes quanto engenharia de materiais ou fisiologia intracelular, porém, talvez não seja o melhor caminho para levar estudantes a dominarem o conhecimento necessário para usar a estatística de uma forma prática na interpretação de dados.

A estatística, da mesma forma que armas de fogo e revistas sobre sexo, pode ser uma ferramenta para promover identidade cultural. Entretanto, acreditamos que o uso da estatística como uma ferramenta de análise de dados e como meio de comunicação entre pesquisadores é o melhor ponto de partida. Se você deseja uma abordagem reducionista, leia Winer *et al.* (1991) ou Harris (1975). Se você quer usar a estatística como uma forma de distingui-lo entre seus pares na academia, pode ler capítulos individuais de qualquer livro-texto de estatística ou, melhor ainda, o texto inteiro de Dytham (1999). Mas se deseja uma alternativa a estas opções, este é o livro correto para você.

Talvez você não precise realmente deste livro. Se você responder "sim" para todas as questões que aparecem na TABELA 1, você domina os principais conceitos necessários para planejar pesquisas e pode se imergir na malha delicada da matemática. Infelizmente, a maioria das pessoas que respondem "sim" para todas as questões da tabela 1 simplesmente o fazem por não serem capazes de perceber o quanto desconhecem.

Mais perigosos ainda, são os que acreditam que a matemática pode suprir a falta de conhecimento dos conceitos referidos na tabela 1. Se o pesquisador não compreendeu bem estes conceitos, nenhuma quantidade de fórmulas tediosas resolvidas à mão, ou em miraculosos programas de computadores e nem mesmo um monte de teoremas matemáticos pode tornar o seu trabalho útil. Este livro não pode fazê-lo competente em todos os aspectos abordados na tabela 1. Na verdade, poderíamos escrever um livro inteiro a respeito de cada um. Concordamos inteiramente com Harris (1975), quando diz que "ainda não encontrou alguém que tenha adquirido domínio em qualquer área da estatística sem ter realizado muitas análises com dados reais – preferivelmente dados realmente importantes para esta pessoa". Contudo, podemos passar para os leitores uma introdução aos conceitos.

Um dos problemas com os livros de estatística é que eles foram escritos por estatísticos. Um estatístico é aquele tipo de pessoa que enxerga o mundo em termos de abstrações matemáticas e que se sente confortável com conceitos que não têm contrapartida no mundo real (Guttman 1985), embora tenham descoberto há muito tempo que o domínio da estatística só vem após uma base em amostragem e inspeção dos dados brutos (p.ex. Deming 1975, Tukey 1980).

TABELA 1. UM GUIA PARA PLANEJAR SEU ESTUDO

Você está preparado para começar a coletar os dados de sua pesquisa?

Se você responder "sim" para todas as perguntas abaixo, então não precisa ler este livro

01. Você decidiu qual é o objeto de seu estudo (variável dependente)?

02. Sua variável dependente pode ser medida objetivamente e você perguntou a outros pesquisadores se eles consideram sua medida "objetiva"?

03. Você consultou os outros membros de sua equipe de pesquisa para se certificar que todos têm os mesmos propósitos?

04. Você esboçou um diagrama de fluxo que mostra quais variáveis influenciam a variável dependente e as relações entre as variáveis independentes?

05. Todos os membros de sua equipe estão coletando dados na mesma escala e nos mesmos lugares, de forma que seja possível integrar os dados ao final do estudo?

06. Você decidiu qual é o seu universo de interesse e todos os membros da equipe concordam com isso?

07. Você desenhou um mapa ou um diagrama conceitual que mostra onde/quando suas amostras serão feitas em relação ao seu universo de interesse?

08. Você desenhou gráficos de pontos hipotéticos, mostrando o número de observações independentes, a variabilidade nos dados e a magnitude dos efeitos que você espera encontrar?

09. Você otimizou o tamanho, forma, orientação e distribuição de suas unidades amostrais de tal forma que a variabilidade na variável dependente seja principalmente devida às variáveis independentes que você está estudando?

10. Sua amostragem está na mesma escala que sua(s) questão(ões)?

11. Você decidiu se está interessado em efeitos diretos, indiretos ou em efeitos gerais?

12. Você decidiu se seus resultados serão usados para determinar se existe um efeito, para determinar a magnitude do efeito nas condições presentes, ou se para predizer o que acontecerá se as condições mudarem?

13. Você se sente confiante de que sua formação em estatística é suficiente para torná-lo capaz de realizar todas as operações mencionadas acima?

14. Você se sente confiante de que sua formação estatística preparou você para escolher a análise apropriada para responder sua questão, antes que você comece a sua coleta de dados?

15. Se você respondeu sim para a questão 14, você consultou um estatístico, mostrando a ele os resultados de todas as operações mencionadas acima, para se certificar que você não está apenas enganando a si próprio?

16. O uso das estatísticas e análises que você escolheu o ajudará a se comunicar com sua audiência?

Nosso curso basicamente segue as recomendações da *"American Statistical Association/Mathematical Association of America joint curriculum committee"* (veja Moore 1997: Fig. 1). A maior diferença é que nós ensinamos os conceitos usando gráficos simples e, quando necessário, analogias. Enquanto os cursos regulares de estatística gastam um dia explicando os conceitos e 9 dias afogando os estudantes em matemática, nosso curso emprega 10 dias na exploração de conceitos em relação às técnicas e análises mais freqüentes na literatura ecológica, e deixa a matemática para cursos subseqüentes ou estudo individual. Esta abordagem funciona bem tanto para estudantes que nunca tiveram um curso de estatística, quanto para

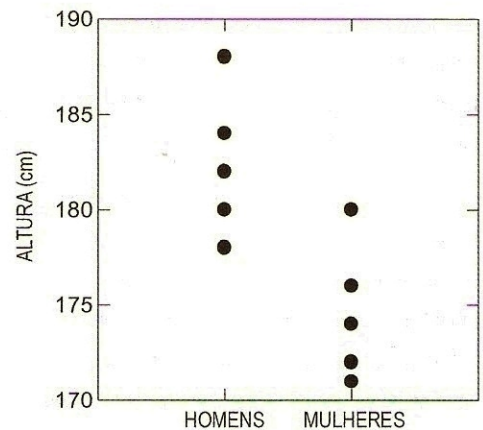
estudantes e profissionais que já tiveram cursos avançados. Alguns dos nossos alunos mais entusiastas são responsáveis por ministrar cursos de estatística para estudantes universitários. A maioria manifestou vontade de aprofundar seu conhecimento de matemática e muitos disseram que gostariam de repetir os cursos de estatística que fizeram anteriormente. Este efeito é muito diferente do que o provocado pela maioria dos cursos tradicionais de estatística, que tendem justamente a aumentar a fobia à matemática, geralmente presente nos biólogos. Em defesa dos biólogos, lembramos que há muitas formas de inteligência e a proficiência em matemática é apenas uma delas (Goleman 1995).

Enfatizamos bastante a comunicação. Um dos nossos problemas com técnicas estatísticas é que cada disciplina tem suas escalas de amostragem e tipos de análises características. As agências financiadoras de pesquisa vêm pressionando cada vez mais os pesquisadores a submeter projetos integrados, que são usualmente chamados de interdisciplinares ou multidisciplinares. Nestes projetos, o líder precisa reunir as contribuições de diferentes pesquisadores, preocupando-se em organizá-las de modo que a proposta final tenha coerência e unidade. Usualmente, o líder não tem experiência nos diferentes campos de estudo do projeto "integrado", e aceita os delineamentos amostrais apresentados por seus colegas. Isto faz com que as equipes de pesquisa trabalhem no mesmo local, freqüentemente lado a lado, o quê não implica, necessariamente, em uma análise integrada dos dados. Algumas vezes, a criação de bancos de dados, ao invés da interpretação deles, passa a ser o objetivo principal dos projetos (Hale 1999). Em conseqüência, a maioria dos estudos é publicada independentemente em revistas especializadas e os resultados integrados, que as agências financiadoras esperavam, não aparecem.

Muitos cientistas são pessoas com dificuldades no trato social, que não gostam de serem vistos por seus pares como sendo diferentes. Neste caso, "pares" significa outros cientistas que militam na mesma disciplina e não os colegas do grupo "integrado". Eles vêem a estatística como um troféu cultural ao invés de um meio de comunicação de informações objetivas. Salsburg (1985) vai a ponto de se referir à estatística como uma religião. O líder e os membros mais esclarecidos do grupo freqüentemente se deparam com a difícil tarefa de convencer os demais a ajustarem seus esquemas de amostragem em função da questão global que está sendo estudada e não às padronizações de cada disciplina. É provável que o líder do grupo seja competente política e socialmente, mas com pouco preparo em matemática ou estatística. Provavelmente, ele/ela tem apenas uma vaga idéia de como integrar os diversos protocolos de amostragem e menos idéia ainda das conseqüências matemáticas de não integrá-los.

Uma solução simples seria o líder do grupo fazer cursos avançados de matemática e estatística e simultaneamente, manter seus contatos políticos e postergar o início do projeto até quando se sentisse matematicamente competente. Isto seria o mesmo que exigir que um operário industrial ganhe a vida pescando: pode funcionar na teoria, mas não vai parecer certo quando a família começar a passar fome. Este livro foi elaborado para fornecer aos líderes e participantes de projetos integrados, informações suficientes para que eles entendam a necessidade e as limitações de protocolos de amostragens efetivos, sem tentar transformá-los em matemáticos profissionais. Todos os conceitos são apresentados com o mínimo de matemática: só fornecemos o necessário para que o pesquisador seja capaz de se comunicar com um estatístico quando julgar oportuno o aconselhamento especializado e para entender o jargão que outros membros da equipe tenham memorizado.

Assumimos que os leitores podem interpretar gráficos simples como o da figura 1, que descreve algumas medidas de altura de um grupo de homens e mulheres. Vamos tentar mostrar que os testes estatísticos mais úteis produzem resultados que podem ser interpretados em termos de gráficos simples como este. Delineamento amostral e cálculos estatísticos não são necessários se as informações originais puderem ser diretamente expressas em gráficos bidimensionais. No entanto, acreditamos que os conceitos por detrás de muitas das análises estatísticas e os resultados que elas produzem, podem ser ensinados graficamente e escrevemos este livro para tentar convencê-los disto.

FIGURA 1

O que é delineamento amostral?

[Delinear uma amostragem é coletar os dados de forma que você tenha uma boa chance de tomar uma boa decisão.] Em grande parte, depende apenas do bom senso, mas ainda assim, nos capítulos seguintes, vamos mostrar alguns exemplos de como modelos simples podem ajudar a revelar padrões que a princípio estavam escondidos.

O delineamento amostral pode ser tão crítico, que pequenas diferenças nos procedimentos de amostragem fazem com que ele seja apropriado ou não para responder determinada questão. O pior é que muitas vezes não nos damos conta disto. Considere, por exemplo, a estória a seguir. Uma espécie de sagüi ocorre apenas em uma reserva e acredita-se que sua população está em declínio. Algumas pessoas sugeriram que a espécie ocorre em maior densidade nas partes da reserva com maior densidade de árvores e isto tem implicações em termos das ações de manejo necessárias para a conservação da espécie em questão. As autoridades responsáveis pela proteção da vida silvestre encomendam um estudo de 2 anos para determinar o quanto as densidades do sagüi estão associadas com as densidades de árvores e se a população do tal sagüi está realmente em declínio. O biólogo A é contratado e decide fazer uma contagem dos sagüis e das árvores seguindo parcelas alongadas, que atravessam a área em estudo. Ele faz uma contagem no primeiro ano e uma segunda no ano seguinte, para comparar a densidade de sagüis entre anos. Os diagramas da figura 2 mostram os dados crus obtidos pelo biólogo A em seus levantamentos, onde cada sagüi aparece simbolizado por um "x" e cada árvore por uma bola, e os gráficos da figura 3 mostram como ele apresentou seus dados às autoridades.

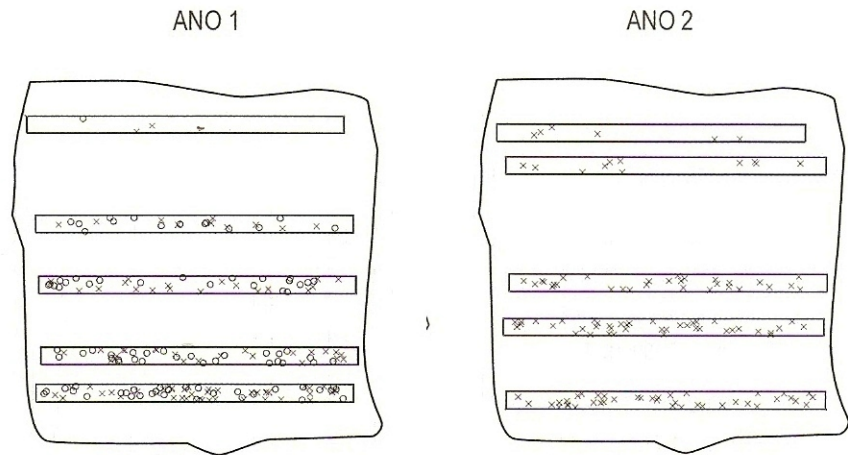
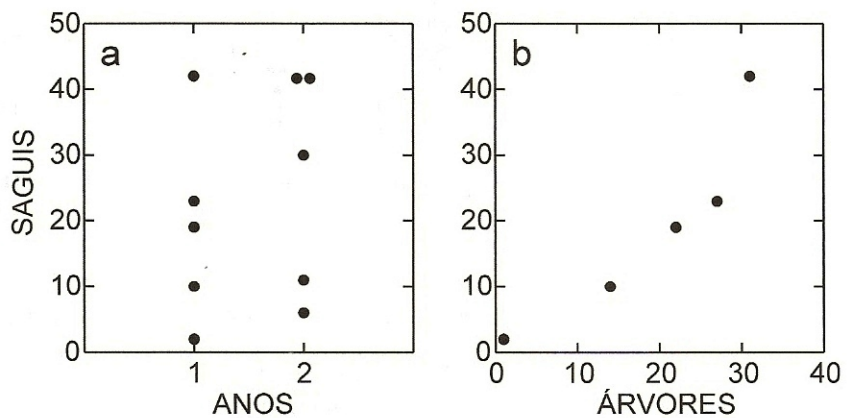
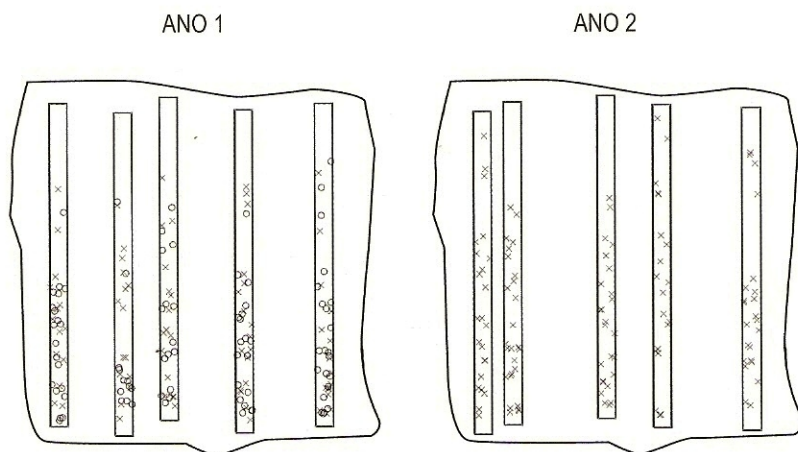
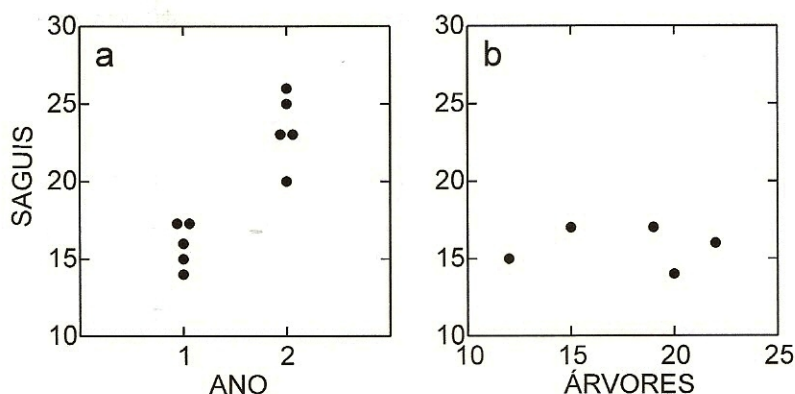
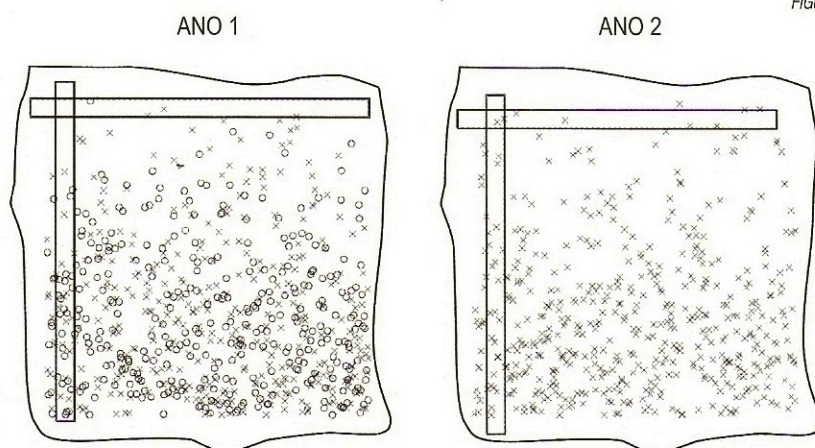
FIGURA 2

Na figura 3a não vemos nenhuma diferença convincente nas densidades do sagüi entre os anos de estudo. Contudo, na figura 3b podemos observar uma forte tendência das densidades do animal serem maiores onde as densidades de árvores são maiores.

FIGURA 3

Vamos imaginar que uma organização conservacionista suspeitasse das intenções do governo e contratasse a bióloga B para fazer um estudo independente endereçado em responder as mesmas questões. Ela usa um delineamento

amostral quase idêntico, exceto que alinha suas parcelas numa direção perpendicular em relação aos transectos do biólogo A (figura 4).

Seus resultados, apresentados na figura 5, indicam uma diferença convincente nas densidades do animal entre os anos, mas uma relação fraca ou inexistente entre as densidades do animal e das plantas (figura 5b).

Os dois biólogos chegaram a conclusões completamente opostas. Contudo, a diferença no delineamento amostral foi apenas a direção das parcelas. Na verdade, os dois biólogos estudaram os mesmos dados. Criamos este exemplo superpondo as parcelas alongadas sobre o mesmo diagrama, mostrado na figura (6), sendo a direção das parcelas a única diferença que atribuímos entre os delineamentos amostrais dos dois biólogos.

Nenhum dos biólogos foi mais correto do que o outro. O delineamento do biólogo A foi superior para detectar a relação entre plantas e animais do que o usado pela bióloga B, mas um esquema diferente de amostragem, baseado em parcelas quadradas ou circulares poderia ser igualmente efetivo. O delineamento usado pela bióloga B foi superior para detectar diferenças entre anos e o uso de parcelas circulares ou quadradas ou qualquer outro formato menos alongado não seria adequado para responder esta questão.

FIGURA 6

Em situações como esta, onde há um forte gradiente nas densidades através da área a ser amostrada, a orientação, forma e tamanho das unidades amostrais irão determinar as questões que podem ser respondidas. Caughley e Sinclair (1994: Capítulo 12) fornece exemplos para vertebrados, Stern (1998), para plantas, e Johnson *et al.* (1999) discutem como amostrar características de habitats. Krebs (1998) oferece uma discussão sobre aspectos gerais de forma e tamanho das unidades amostrais.

Mesmo quando não há gradientes fortes, como no nosso exemplo, sempre há uma escala em que os organismos estão agrupados. Não podemos nos estender muito em como selecionar as unidades amostrais neste livro. Entretanto, a menos que a forma e o tamanho das unidades amostrais seja apropriada para uma determinada questão, nenhuma das técnicas estatísticas, nem as mais sofisticadas, que discutiremos nos próximos capítulos, terão utilidade para revelar padrões da natureza ou comunicar resultados de pesquisa.

[Tukey (1980) enfatizou que "Encontrar a pergunta certa é freqüentemente mais importante do que encontrar a resposta certa".] Não podemos ajudar muito os leitores neste aspecto crucial, já que formular questões realmente interessantes envolve *insight*, experiência e curiosidade. Mas não basta que as perguntas sejam interessantes. Elas precisam ser "respondíveis", no sentido de que devem dar origem a hipóteses que possam ser refutadas (veja capítulo 5). Perguntas do tipo "existe vida depois da morte?" são evidentemente interessantíssimas, mas desafiam a mente humana a derivar hipóteses refutáveis. Perguntas como estas se situam além da ciência e da física atual, são metafísicas e a biologia é repleta de questões metafísicas. Formular perguntas interessantes e respondíveis é arte. Mesmo assim recomendamos a leitura de Tukey (1980) e Guttman (1985), para aqueles que desejam uma orientação adicional neste tópico.

O delineamento amostral diz respeito à compreensão de conceitos que são importantes a cada passo no processo, desde o planejamento, execução, análise e publicação. Este livro não é como os livros regulares de estatística, embora não traga nada que não possa ser encontrado nos livros regulares, se souberem onde procurar. Também não é

um manual de um programa de computador (veja Dytham 1999 para uma introdução à computação estatística). Este livro cobre apenas os princípios mais gerais que os pesquisadores precisam entender para usar as estatísticas convencionais de forma inteligente.

Não tentamos ser completos. Na verdade, acreditamos que este livro será tanto mais útil quanto mais assuntos deixarmos de fora dele. Nós sofremos cada vez que decidimos excluir um ponto importante ou detalhes interessantes, mas nossa experiência indica que muita informação retarda o aprendizado e a compreensão de conceitos. Uma vez armado destes conceitos o pesquisador pode rapidamente descobrir os detalhes. Não cobrimos os tópicos que a maioria dos pesquisadores parecem entender. Escolhemos aqueles que os pesquisadores e estudantes têm dificuldade em entender em seus cursos regulares de estatística, e que causam a maioria dos problemas de comunicação entre pesquisadores. Além disso, fornecemos o mínimo de referências possível, já que a maior parte do que veremos aparece em *qualquer* bom livro-texto de estatística, embora quase sempre tão escondido que a maioria dos leitores simplesmente não pode encontrar. Muitos de vocês podem simplesmente ignorar a maioria das citações, já que elas podem distrair a atenção dos conceitos mais importantes. Contudo, os professores estão convidados a se referirem à literatura original, ao invés de nossos sumários necessariamente breves e incompletos. A literatura que citamos é fortemente enviesada em direção a abordagens conceituais e filosóficas, ao invés de em técnicas matemáticas, exceto talvez, na seção sobre estatística multivariada, devido à complexidade especial deste tópico.

O que esperamos que você obtenha deste livro

Há conceitos importantes que constituem a base da maior parte da comunicação científica. Para comunicar estes conceitos preferimos usar exemplos mas, ocasionalmente, tivemos de usar alguns cálculos. Recomendamos que os leitores não se acanhem diante deles, porque estão sendo usados para passar conceitos importantes. A ordem em que apresentamos os capítulos neste livro é ligeiramente diferente da ordem em que os apresentamos em nosso curso, já que em sala de aula, a presença do professor com todo seu carisma permite uma organização menos metódica da matéria a ser apresentada. Talvez aqueles leitores que também sejam professores de cursos de estatística, ou de delineamento amostral, queiram começar com o capítulo 13 "Dicas para professores".

[*www.editoraplanta.com.br*]

CAPÍTULO 2. "Fluxogramas e questões científicas". Através de todo o livro iremos nos referir a diagramas que descrevem hipóteses ecológicas. Nenhum teste estatístico pode ser interpretado, sem que esteja relacionado a um fluxograma, embora talvez os leitores não possam entender isto até o final do capítulo 10. Neste estágio, fluxogramas são importantes para forçá-lo a ser explícito a respeito de seus objetivos e ajudá-lo a começar a compreender a diferença entre as variáveis que causam efeito (variáveis independentes) e variáveis que são afetadas (variáveis dependentes). Ainda no capítulo 10, há, também, uma breve discussão da importância da escala em ecologia, o que é relacionado a problemas discutidos já no capítulo 2.

CAPÍTULO 3. "Descrevendo a natureza". Este é o capítulo mais parecido com livros convencionais de estatística. Entretanto, não se preocupem muito com as fórmulas. O importante é que você compreenda que alguns conceitos como "desvio padrão" podem ser visualizados em gráficos enquanto outros, como "variância", não são fáceis de se visualizar. Nós exploramos o "erro padrão" para ensinar alguns outros conceitos em nosso curso (*veja "Dicas para professores"*). De qualquer forma, não há necessidade de se memorizar fórmulas, também aquelas técnicas que envolverem conceitos difíceis de serem visualizados em gráficos serão explicadas por analogias nos capítulos subseqüentes. O capítulo 3 é mais importante como uma introdução ao conceito de "desvio" e por fornecer uma base que permita ao leitor interpretar a literatura. Esperamos poder convencê-los de que a maioria da estatística "descritiva" obscurece os dados mais do que os revelam, e esperamos apresentar (reapresentar) a vocês a ferramenta de comunicação científica mais importante, o gráfico de dispersão (gráfico de pontos ou *"scatterplot"*).

CAPÍTULO 4. "Quanta evidência é necessária?" Este capítulo apresenta a relação entre a força da inferência e o número de pontos em um gráfico de dispersão. Discute também sobre informações que não aumentam a força de nossas inferências, o que ficou conhecido entre os ecólogos como "pseudo-repetição" (no Inglês *"pseudoreplication"*). Aqueles que freqüentam a literatura científica precisam aprender reconhecer, ou pelo menos suspeitar de pseudo-repetições espaciais, temporais, filogenéticas e técnicas. Precisam entender que nenhuma observação é intrinsecamente válida ou uma pseudo-repetição. Isto depende inteiramente da questão que está sendo formulada.

CAPÍTULO 5. "Quando *improvável* significa *bem possível"*. Aqui apresentamos a filosofia popperiana, que está por trás da maioria das correntes de pensamento predominantes na estatística. Não é possível entender a estrutura da maioria dos testes estatísticos, a menos que se entenda os conceitos sob uma perspectiva popperiana. Ela é também a base para "dendrogramas de decisões" e outros procedimentos científicos que sequer envolvem cálculos matemáticos. Ciência sem filosofia é uma coisa perigosa, e o mesmo se dá com a estatística.

CAPÍTULO 6, "Evitando riscos em comparações simples", introduz a análise de variância simples (ANOVA) para fatores categóricos. Aqui, pela primeira vez neste livro, discutiremos explicitamente erros do tipo II, que, embora freqüentemente sejam mais importantes do que erros do tipo I, raramente são considerados nos testes estatísticos. Nossa discussão sobre erros do tipo II é breve, assim, se você não gostar das implicações de se cometer estes erros, recomendamos a leitura de alguns dos trabalhos que citamos neste capítulo. Embora a ANOVA de um fator seja apresentada como uma forma de evitar a acumulação de riscos, o conceito mais importante a ser assimilado é simplesmente a possibilidade de repartir a variabilidade entre o fator atuante e o resíduo. Os leitores precisam compreender este conceito, ou não serão capazes de entender qualquer técnica estatística usual. Por este motivo, procuramos apresentá-lo em gráficos simples, e rogamos que se detenham sobre eles o tempo necessário para absorvê-lo completamente.

CAPÍTULO 7, "Análises para um mundo com todos as suas tonalidades" trata de uma "ANOVA" com fatores contínuos, que geralmente é chamada de regressão. Neste capítulo vocês deverão aprender que o mundo consiste de variáveis contínuas e que converter variáveis contínuas em categorias quase sempre é contra-produtivo e freqüentemente, enganador. Entretanto, aprenderão que o conceito de uma única partição da variabilidade nos dados em suas fontes de variação, se aplica tanto para as variáveis categóricas quanto para as contínuas.

Na verdade, a ANOVA de fatores categóricos, que é referida nos livros de estatística como "Análise de variância", nada mais é do que um caso especial de regressão. Há outras maneiras de se atacar questões de apenas um fator, mas, para não quebrar a seqüência lógica, a deixaremos para o capítulo 11. Isto não seria necessário dentro de sala de aula (veja "Dicas para professores").

[*www.editoraplanta.com.br*]

CAPÍTULO 8. "Problemas do mundo real: mais do que um fator." Este título é uma pequena pretensão de nossa parte. Na verdade, este ainda não é o "mundo real", e o título apenas mostra que modelos mais complexos usualmente são requeridos para começar a responder questões ecológicas. Infelizmente, muitos pesquisadores acreditam que estas análises modelam situações do mundo real. As análises aqui empregadas ainda são baseadas em modelos lineares simples e aditivos, que permitem uma partição única de efeitos entre os fatores. Não passe para os capítulos seguintes enquanto não estiver confiante de que compreendeu o conceito de alocação das variâncias entre diferentes fatores, e o conceito de interação entre fatores.

CAPÍTULO 9. "Quais variáveis analisar?" Provavelmente não respondemos esta questão neste capítulo. A engenhosidade e a experiência necessária para a seleção ótima das variáveis são parte da arte do naturalista e não podem ser ensinadas nos livros. Contudo, usamos os conceitos aprendidos nos capítulos e exemplos anteriores para mostrar aos leitores o que *não* devem fazer para selecionar variáveis.

CAPÍTULO 10, "Amarrando as coisas", continua a utilizar modelos lineares aditivos para modelar o mundo. Porém, aqui mostramos que, exceto em situações extremamente simples e freqüentemente triviais, não existe uma variabilidade única que pode ser atribuída a cada fator. Este capítulo deve comunicar, sem sombra de dúvida, a importância dos diagramas que foram descritos no capítulo 2. O leitor deve compreender as diferenças entre efeitos diretos, indiretos e gerais, e porque nenhum teste estatístico pode ser interpretado sem que esteja relacionado com um fluxograma.

CAPÍTULO 11. "Endireitando o mundo: transformações e outros truques." Pela lógica, este tópico deveria vir em seguida ao capítulo 7, como fazemos em sala de aula. Entretanto, em um livro isto poderia distrair o leitor da discussão sobre as técnicas de alocação de variâncias, que é a base de cerca de 90% da estatística encontrada na literatura científica. As técnicas descritas neste capítulo usam uma variedade de métodos para estimar os parâmetros que descrevem nossos modelos matematicamente, e eles freqüentemente podem lidar com curvas

complexas. Porém, elas têm sido pouco usadas, porque geralmente não permitem a alocação da variância entre as fontes e não podem ser usados para se determinar a importância relativa de cada fator, a não ser que lancemos mão de simulações complicadas usando computadores, ou nos contentemos em dividir os fatores em categorias de "significantes" e "não significantes".

Talvez, depois de compreender todas as limitações das técnicas de alocação de variâncias, os pesquisadores mais avançados comecem a considerar seriamente o emprego de técnicas de simulações para refinar seus modelos.

CAPÍTULO 12. "Análise multivariada." Diz-se que "os tolos correm por caminhos onde os prudentes vão passo-a-passo". Infelizmente, os inexperientes também correm para a estatística multivariada. Muitos estudantes (e seus orientadores) pensam que a estatística multivariada é um remédio para todos os males. Freqüentemente, estudantes podem gerar enormes matrizes de dados que impressionam a maioria das pessoas, mas que em muitos casos não têm repetições suficientes nem para o exame de um único fator. Entretanto, se um estudante intui um padrão na relação de diferentes variáveis, através do exame de gráficos ou tabelas, e busca este padrão usando técnicas multivariadas, é possível que os resultados não sejam apenas um artefato estatístico.

Mesmo que os estudantes não pretendam utilizar técnicas multivariadas, precisam compreender seus princípios gerais, para serem capazes de entender a literatura. Tentamos mostrar os princípios gerais sem revisar a matemática envolvida, e apontar as dificuldades mais comuns. Os leitores deverão ser capazes de pelo menos conceituar a relação entre dimensões "fantasmas" e gradientes reais antes de terminar com esta sessão.

CAPÍTULO 13 "Como escrever melhor de trás para frente." Uma das partes mais difíceis da pesquisa é apresentar suas conclusões de maneira lógica de modo que os leitores possam ver a relação entre a questão, os gráficos e as conclusões. O último capítulo impresso fornece sugestões de como fazer isto.

[Para aqueles que irão adotar este livro em sala de aula, o suplemento "Dicas para Professores" apresenta a seqüência de aulas que funcionou melhor em nossa experiência e também fornecemos exemplos que podem ser usados nos exercícios de classe.]

[*www.editoraplanta.com.br*]

CAPÍTULO 2

[Ciência é uma arte, e arte é comunicação]

Fluxogramas e questões científicas

Ciência é uma arte, e arte é comunicação. Um pintor vê uma paisagem e determina o que ele quer retratar (p. ex. harmonia, quietude, grandiosidade) e representa esta qualidade essencial em duas dimensões, usando cor, textura e forma. Dependendo da escola a que o artista pertença, ele pode querer transmitir algo a respeito da paisagem ou de si mesmo, ou ambos. Um cientista faz quase a mesma coisa. O ecólogo, olhando para a paisagem, pode pensar em reduzi-la a uma qualidade essencial (p. ex. competição, mutualismo, restrições físicas, metabolismo) e representa esta qualidade em duas dimensões, usando palavras, gráficos e fórmulas matemáticas. Muitas vezes o cientista acredita que sua representação da realidade é "objetiva" e a única que uma pessoa racional poderia fazer. Contudo, cedo ou tarde ele aprende, através de sua experiência pessoal ou pelo estudo da história da ciência, que sua representação é apenas parcial e que está distorcida pelos filtros de sua cultura e de sua época. Assim, ele considera alternativas e tenta calcular a probabilidade de estar errado. Este processo é formalizado na matemática da estatística inferencial. Em ciência, espera-se que o autor esteja comunicando mais sobre a paisagem do que a respeito de si mesmo.

A principal contribuição do líder de um grupo integrado de pesquisa é a elaboração de um diagrama bi-dimensional do sistema que está sendo investigado. Vamos chamar este desenho de "fluxograma". Contudo, um engenheiro chamaria isto de "análise de sistemas", enquanto um psicólogo falaria em "modelos causais". Ecólogos freqüentemente fazem "análises de caminhos" (*Path Analyses*). Os detalhes desses métodos não importam no momento. O importante é entender que diferentes pessoas enfrentando problemas complexos fazem uso de técnicas similares e essas técnicas não são propriedade de uma ou outra disciplina. Há uma arte para fazer fluxogramas e algumas regras básicas, mas apenas a prática leva à competência. Nós descobrimos que dos modelos complicados, os fluxogramas são os mais fáceis de serem entendidos pelos estudantes e, portanto, é um bom lugar para se começar a tratar da complexidade da natureza. Alguns estudantes mais avançados e professores de cursos de bioestatística podem querer procurar em Higashi e Burns (1991) outras maneiras de conectar elementos dos ecossistemas.

Comece por decidir o que você está estudando. Isto precisa ser alguma coisa mensurável. Conceitos complexos como "qualidade ambiental", "estado de conservação" e "justiça social" não têm dimensões, ou pelo menos não têm dimensões que podem ser reconhecidas pela maioria das pessoas e por isso não serão úteis. O que é que realmente se deseja medir? Qualidade ambiental poderia significar condições que propiciem longas expectativas de vida para seres humanos, condições que propiciem a seres humanos uma ampla variedade de atividades ao ar livre, condições que permitam a perpetuação de comunidades de animais ou plantas que existiam no local quando as populações humanas eram pequenas, ou qualquer outra de uma multidão de condições que qualquer pessoa considere indicativa de "qualidade".

É preferível envolver toda a equipe no processo de decisão sobre quais são as questões do estudo, mesmo quando aparentemente as questões tenham sido determinadas pela agência financiadora. Esta parte do estudo é usualmente a mais difícil e, por essa razão, é evitada na maioria das propostas. Não há questões implícitas ou óbvias em um projeto de pesquisa. Se a questão não foi explicitamente colocada, significa que o líder do projeto está confuso, ou é incompetente, ou pior ainda, desonesto. São palavras duras, mas a conseqüência de questões vagas é o desperdício de tempo, dinheiro e credibilidade dos cientistas. Hobbs, (1988) apresenta um fluxograma engraçado, mas ao mesmo tempo trágico, que representa a contribuição de pesquisas ecológicas na tomada de decisões. Cada dólar desperdiçado em uma pesquisa ruim, poderia ser usado para salvar vidas de crianças carentes. Temos trabalhado em países e regiões pobres e admitimos que nos tornamos intolerantes com esforços de pesquisa desperdiçados. Há muitos textos sobre processos de decisão disponíveis nas livrarias e não vamos nos alongar neste assunto. Uma boa introdução a esta discussão pode ser encontrada em Tukey (1960, 1980) e no capítulo 1 de Caughley e Sinclair (1994). Uma discussão mais "estatística" dos problemas na seleção das questões e medidas em estudos integrados pode ser encontrada em Osenberg *et al.* (1999). Vamos assumir que a questão ou a variável a ser investigada já foi determinada e focalizar o problema de decidir o que pode afetar esta variável e de que maneira.

Construindo a hipótese inicial

Você pode começar com a premissa de que qualquer coisa é conectada com tudo o mais e tentar colocar tudo em seu modelo. Isto é análogo a um escultor querer colocar uma montanha inteira sobre o seu pedestal - uma perda de tempo. Por outro lado, você pode montar um sistema tão simples que ele não tenha mais semelhança com o mundo real e, novamente, desperdiçar esforços. Nosso modelo deve ser simples o suficiente para ser manejável, mas complexo o suficiente para capturar a idéia central do problema. Esta é a arte do cientista. Uma boa discussão sobre construção de modelos pode ser encontrada em Starfield e Bleloch (1991). Os leitores que se sentem confortáveis com uma matemática um pouco mais complexa podem

[consultar Burnham e Anderson (1998). Contudo, estamos interessados em um nível muito mais geral do que estes autores e **a construção de modelos é uma arte que vem com a experiência.** Não se envergonhe se seus primeiros modelos não parecerem bons. Mesmo um modelo ruim é útil, nem que seja para mostrar as limitações do pesquisador ou de seus dados.]

Neste capítulo, e no resto deste livro, iremos tratar de modelos e suas premissas. Por vezes, a construção de modelos aparenta ser um processo técnico. Entretanto, modelos envolvem conceitos filosóficos mais complexos do que a matemática. Apresentamos duas citações de Allen (1998), que refletem também o nosso ponto de vista filosófico. "Modelos, que não passam de réplicas dos sistemas reais, não podem ser a representação fiel dos sistemas em todas as situações. Portanto, todos os modelos são errôneos, no sentido de diferirem das observações. Obviamente, a noção de certo e errado não serve para nada em modelagem, e dizer que um modelo, que tenha sido construído com lógica e consistência, é acertado ou errôneo, não vem ao caso... Todas as premissas são falsas em algum nível, portanto, a correção das premissas está fora das possibilidades, e também não vem ao caso". A ciência não diz respeito à obtenção de respostas para tudo; refere-se a encontrar situações onde se possa ir adiante com uma premissa – e ainda conseguir que o modelo tenha alguma utilidade. Se você espera respostas absolutamente certas para todas as questões, sugerimos que abandone este livro e procure suas respostas na teologia.

FIGURA 7

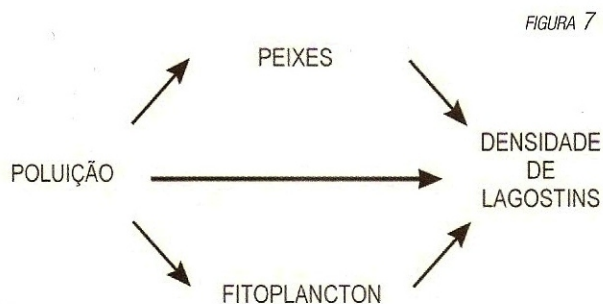
PEIXES

POLUIÇÃO → DENSIDADE DE LAGOSTINS

FITOPLANCTON

Vamos começar com um exemplo simples, relacionado com uma questão ecológica (figura 7). A mesma lógica pode ser aplicada a estudos em outras escalas, como em experimentos laboratoriais sobre fisiologia ou questões sobre padrões biogeográficos em escalas continentais.

Este fluxograma é, obviamente, uma simplificação tão grande, que não permite a análise na presente forma. Mas, ainda assim, vamos usá-lo para discutir os mecanismos de construção de um diagrama. Este fluxograma preliminar é importante porque ilustra a hipótese inicial e as premissas. Você obviamente está interessado na densidade de uma espécie de lagostim, presumivelmente porque ele tem importância comercial ou por estar sendo considerado como ameaçado de extinção, ou alguma outra razão que justifique os gastos. Caso se descubra que os membros da equipe não estão de acordo com a questão geral, você deveria repensar o problema antes de submeter à proposta.

Examinando o fluxograma mostrado na figura 7, descobrimos que a equipe acredita que os fatores mais importantes que estão afetando a densidade do lagostim são o fitoplancton, a poluição do corpo d'água e peixes

predadores. Espera-se que experiência, informação da literatura e bom senso sejam usados para decidir quais os fatores mais importantes a serem estudados. Milhares de outras coisas que potencialmente podem ser importantes, como predação por aves, tipo de substrato e doenças, não foram incluídas. Esta é uma decisão subjetiva e os pesquisadores podem estar errados. Entretanto, eles corajosamente mostraram quais fatores eles acreditam ser prioritários para a pesquisa, e os leitores podem facilmente ver quais fatores foram deixados de fora. Eles serão criticados por muitos biólogos de escritório, que exortarão a necessidade de inclusão de inúmeras outras variáveis que poderiam ter *algum* efeito sobre o lagostim. Este é o preço da honestidade e integridade. Além disso, alguns dos revisores podem até tecer críticas realmente construtivas.

A primeira coisa a notar no fluxograma, é que as setas indicam a direção do efeito. Uma seta com pontas nas duas extremidades indica que cada uma das variáveis influencia a outra. Se não há pontas de flecha no traço que une duas variáveis, indica que elas variam conjuntamente, mas nenhuma afeta diretamente a outra. Embora teoricamente possível, nenhuma destas duas situações é útil para modelagem e o leitor deveria procurar uma terceira variável que explique a relação.

A figura 8 mostra uma correlação, que pode ser substituída pela figura 9. Se duas variáveis têm uma relação causal, como aparece na figura 10, deveríamos ser capazes de incluir um ou mais fatores que explicassem esta causalidade, da forma exemplificada na figura 11.

FIGURA 8

FIGURA 9

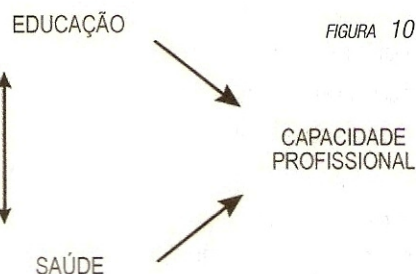
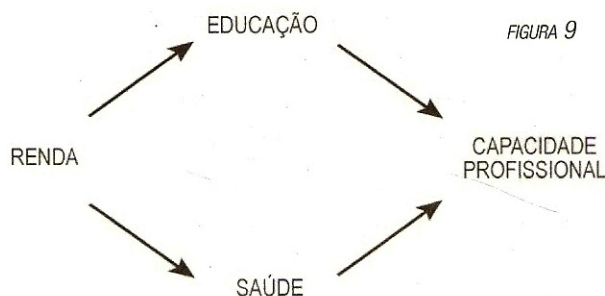
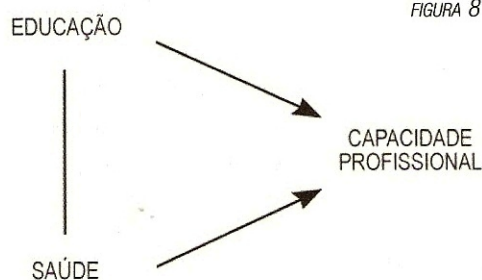
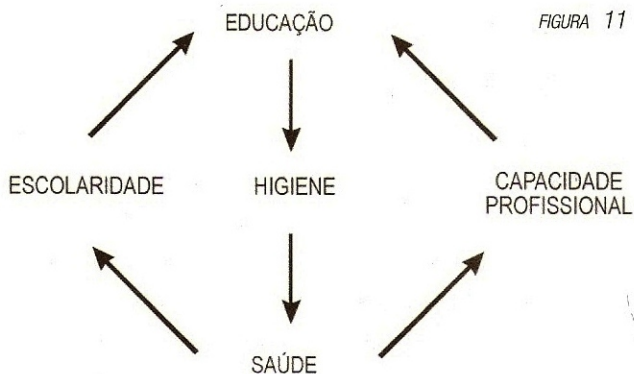
FIGURA 10

Este exercício resultou em fluxogramas que mostram os fatores que julgamos importante para o processo em estudo e o que afeta o que. É claro que todos os modelos apresentados até agora demandam ainda muito trabalho e pode ser difícil admitir que eles são vitais como ponto de partida de estudos científicos. Eles não são "sofisticados" no sentido original da palavra, que era "desnecessariamente complexos", e não contêm aqueles hieróglifos que usualmente associamos com textos científicos, como " $P <$ " ou " $\chi^2 = \ldots$ "

FIGURA 11

Surpreendentemente, estes diagramas são necessários para interpretar a maioria das análises estatísticas. Os estudos científicos investigam uma, ou algumas vezes muitas, das flechas de um fluxograma. Raramente ocorre de os pesquisadores investigarem todas as setas de seus modelos, mas sem o diagrama é difícil enxergar onde o estudo se encaixa dentro de todo o esquema. Veremos, também, que só podemos determinar a validade da maioria das análises estatísticas se soubermos onde elas se encaixam no fluxograma.

Retornando ao exemplo do lagostim, vemos que sua densidade presumivelmente depende de muitas outras variáveis. Isto justifica sua designação de "variável dependente" no jargão estatístico. Entretanto, podemos notar que algumas das outras variáveis em nosso modelo (chamadas de "variáveis independentes") são, de fato, dependentes umas das outras. Isto pode complicar nossas análises. Além disso, algumas das variáveis que afetam diretamente a densidade dos lagostins produzem também efeitos indiretos, porque elas influenciam outras variáveis que afetam a densidade do lagostim. Por exemplo, a poluição afeta diretamente a densidade do lagostim, mas também afeta indiretamente, porque a poluição afeta os peixes, que afetam a densidade do lagostim. Existem técnicas estatísticas para lidar com efeitos indiretos e variáveis independentes quando elas não são realmente independentes umas das outras, como as descritas no capítulo 10, mas, por favor, não pule para este capítulo ainda. Raramente poderemos satisfazer as premissas destas análises e a maioria dos pesquisadores de sua equipe estará focalizada em partes limitadas do seu diagrama. Antes, precisamos considerar os diferentes modos dos pesquisadores estudarem as setas que lhes cabem do diagrama, porque a maioria destes modos não contribui para gerar os dados necessários para a análise estatística do fluxograma global.

Três tipos de estudo

1. ISTO É REAL?

A questão básica, e freqüentemente a única questão de muitos estudos, é verificar se um efeito existe. Isto equivale a perguntar se nós deveríamos mesmo incluir esta seta em nosso diagrama. Pode parecer simples provar que um efeito *existe*, mas nada é tão simples assim. Alguns poucos cientistas calculam a probabilidade de que a seta exista, lançando mão da estatística Bayesiana, e sem dúvida os cientistas adeptos da estatística Bayesiana têm facilidade de se comunicar com políticos e com o público em geral. Albert (1997) nos dá uma explanação fácil de ser digerida de como a estatística Bayesiana pode ser usada para responder questões simples. Entretanto, a estatística Bayesiana não é fácil de se entender ou calcular (Moore 1997), e Guttman (1985) refere-se a ela como "uma cura pior do que a doença". A maioria dos testes estatísticos, dos textos de estatística e dos programas estatísticos para computadores são baseados em técnicas freqüentistas e calculam a probabilidade de a seta *não* existir em nossos fluxogramas. Isto está longe do que o senso comum indicaria como o procedimento lógico e não é o tipo de probabilidade com que a maioria das pessoas se sente à vontade. No entanto, a maioria dos testes estatísticos que os membros de sua equipe irão usar será baseada em técnicas freqüentistas (*i.e.* na filosofia Popperiana). Portanto, se o leitor não é familiarizado com a filosofia Popperiana, deveria ler o capítulo 5.

É importante entender que freqüentemente os testes estatísticos mais poderosos não respondem a questão "O efeito normalmente existe?", mas sim à questão "Se todas as outras variáveis forem mantidas constantes, a mudança desta variável produz algum efeito?". Estes testes consideram apenas os efeitos diretos, tornando os efeitos indiretos impossíveis de serem acessados, porque constrangem as outras variáveis, tratando-as como constantes. Esta diferença não é trivial. Alguns fatores que não têm efeitos diretos nas variáveis-resposta podem ser muito importantes no mundo real, e algumas variáveis que normalmente têm um pequeno efeito direto, difícil de ser detectado em campo, podem ter grande potencial para a medicina ou agricultura, se seus níveis puderem ser manipulados artificialmente.

A maneira mais convincente de mostrar que efeitos diretos estão atuando é através de um experimento que manipule o sistema, de forma que apenas as variáveis estudadas possam influenciar o resultado. Diz-se que os resultados destes experimentos permitem uma "inferência forte" (*sensu* Platt 1964), já que as variáveis que poderiam estar confundindo os resultados foram eliminadas. Contudo, os resultados obtidos podem não ter muita relevância para o mundo real (p. ex. Carpenter, 1999) e os ecólogos serão sempre capazes de sugerir um ou mais fatores prováveis de desempenhar *algum* efeito (Tukey 1991, Johnson 1999), mesmo que estes efeitos não tenham muita importância nos sistemas não manipulados.

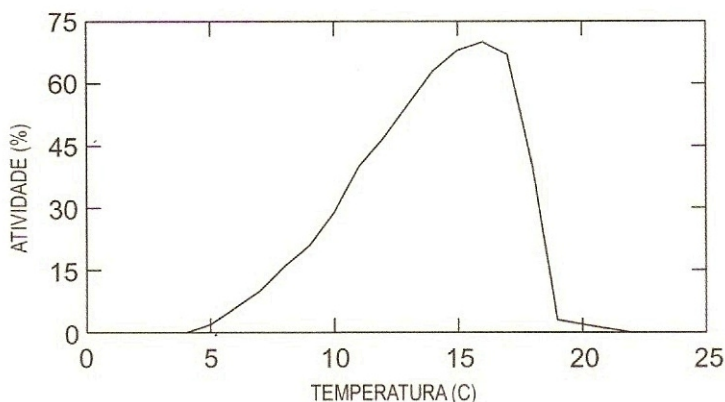
[O líder e os membros do grupo precisam se assegurar de que as questões formuladas por cada um dos pesquisadores da equipe sejam relevantes para a questão geral.]

2. QUAL É A FORMA E MAGNITUDE DO EFEITO?

Descobrir que um efeito é ou não diferente de zero freqüentemente não é muito útil (p. ex. Rosenthal e Rubin 1994). A segunda fase da pesquisa freqüentemente levanta questões a respeito da forma da relação existente ou que se presume existir. A maioria dos modelos é construída sobre relações lineares simples ou linearizadas e nosso fluxograma reflete isto. Poderíamos achar que, para cada peixe predador acrescentado ao sistema, reduzimos a densidade do lagostim em três indivíduos por metro de riacho. Mas, um resultado simples como este é improvável. Muitos estudos ecológicos têm mostrado que o efeito de predadores sobre as densidades das presas não é simplesmente uma função linear de suas densidades. Caughley e Sinclair (1994: Capítulo 11) fornecem muitos exemplos. Lagostins em pequenas densidades podem não ser suficientes para manter populações de peixes. Em altas densidades de lagostins, a população de peixes pode não ser limitada pela disponibilidade de alimento, mas por outros fatores, como, por exemplo, a densidade de seus próprios predadores.

Muitas variáveis que afetam processos biológicos se comportam desta forma, não produzindo efeitos em baixos níveis, efeitos positivos em níveis intermediários e efeitos negativos em níveis elevados. A figura 12 mostra a relação entre a temperatura e a proporção da população em atividade de um inseto hipotético. Pensem quantas relações em seu campo de estudo reagem desta forma.

É importante conhecer a forma da relação entre as variáveis para decidir como vamos incorporá-las em nosso modelo, se devemos mudar as unidades de medida das variáveis (ou seja, transformá-las), ou se vamos decidir investigar apenas uma gama limitada de condições, antes que possamos aplicar a maioria dos testes estatísticos ou técnicas de modelagem. Na verdade, é óbvio que precisamos de informação sobre a forma da relação antes de testar se um efeito existe. Se compararmos os níveis de atividade do inseto a 5°C e a 20°C concluiremos que a temperatura tem pouco ou nenhum efeito, ao passo que se compararmos a atividade entre 5 e 15°C, concluiremos que a temperatura tem um efeito muito forte. Este problema é freqüente quando são aplicados testes para verificar a existência de fenômenos com premissas incorretas quanto ao formato da relação ou quando os pesquisadores não investigaram todos os níveis possíveis do fenômeno. Vamos considerar mais sobre estas questões nos capítulos 7 e 9.

FIGURA 12

3. O QUE ACONTECERÁ SE AS CONDIÇÕES MUDAREM?

Muitos pesquisadores gostariam de responder à questão "Quanto efeito uma variável tem sobre a outra" e este é o terceiro nível do estudo. Infelizmente, usualmente, a resposta é "depende". Para responder a esta questão, devemos investigar os níveis em que esta variável normalmente ocorre no sistema, nas freqüências em que estes níveis ocorrem. Então, normalmente as manipulações experimentais recomendadas para a "inferência forte" não são apropriadas para responder a esta questão. Além disso, quando o efeito de uma variável depende dos níveis de outras variáveis 'independentes", como costuma ocorrer no mundo real, não há uma única resposta correta (*i.e.* a resposta é "depende").

Usualmente, as pessoas que se dedicam à modelagem de sistemas não tentam fazer previsões precisas e não estão interessadas em experimentos de "inferência forte". Elas tentam simular como os sistemas funcionam, usando um computador e relações matemáticas simples. Elas modificam o nível e/ou a variabilidade de cada fator e rodam o modelo centenas ou milhares de vezes para tentar determinar quais as variáveis mais importantes do sistema. Se você tem um modelador em sua equipe, vocês deveriam discutir logo as necessidades dele, porque geralmente os modeladores não podem fazer uso do tipo de dados que a maioria dos pesquisadores coletam. Osenberg e colaboradores (1999) fornecem uma boa discussão sobre as dificuldades enfrentadas em estudos integrados, mesmo quando todos os pesquisadores estão de fato estudando a mesma questão simples. Se a complexidade das estatísticas que os pesquisadores de seu grupo empregam o atordoa, considere delinear o estudo para permitir análises mais simples, ao invés de ficar tentando reparar os erros amostrais por meio do emprego intensivo de matemática.

Vimos que os pesquisadores podem ter questões muito diferentes, mesmo quando estão aparentemente estudando o mesmo fenômeno. Alguns cientistas tentam encontrar uma situação intermediária, coletando dados de uma maneira que permita fazer inferências sobre a probabilidade de um efeito ocorrer, e também obtêm os dados necessários para os modeladores de sistemas. Provavelmente, eles serão criticados por todo mundo.

Não há uma resposta correta, e a ciência tem avançado espetacularmente com pesquisadores diferentes trabalhando com diferentes abordagens. Entretanto, em estudos integrados, tanto o líder quantos os colegas de equipe devem conhecer que tipo de coleta e análise de dados cada membro pretende fazer, e precisam decidir o quanto estes dados serão úteis para os demais. Delineamentos amostrais integrados são freqüentemente os mais eficientes e discutiremos alguns daqui a pouco, mas antes, vamos considerar a escala da questão.

Qual o tamanho do problema?

Muitas das controvérsias na literatura científica ocorreram porque pesquisadores trabalharam sobre o mesmo problema em escalas diferentes. A densidade de uma espécie de planta pode não ser relacionada com características do solo em uma escala de dezenas de metros, embora as características do solo expliquem grande parte da variação na densidade de plantas entre regiões. Mas em uma escala mais ampla ainda, por exemplo, em uma escala continental, pode não haver relação entre a ocorrência de espécies e os tipos de solos. A resposta para a questão de se o solo afeta a distribuição de espécies é "não, sim, não". Se tivéssemos examinado outras escalas, poderíamos ter obtido outras combinações de sim e não. Portanto, não é surpreendente que os leitores de revistas científicas fiquem confusos com os "avanços" da ciência e o público geral ainda mais.

Allen e Starr (1982) apresentam uma série de fotografias de um jogo de futebol que mostra como a escala na qual se observa um sistema determina o que podemos descobrir. Quando visto de um avião, não é possível tirar conclusões a respeito do jogo, embora possamos ver todo o estádio e ter uma boa idéia do comparecimento do público. Mais perto, podemos ver os jogadores, mas ao fundo, a imagem do público nas arquibancadas se mistura à dos jogadores. Na aproximação seguinte, é possível observar a ação de jogadores individuais e o jogo em progresso. Numa escala ainda mais próxima, em um ponto de vista no interior da bola, a fotografia é completamente negra. Os pesquisadores pensam bastante nestes aspectos quando vão assistir a jogos de futebol, mas, surpreendentemente, costumam dar pouca atenção a eles quando planejam seus estudos científicos.

É claro que não há uma escala "correta" para os estudos científicos. Muitos artigos recentes têm discutido a importância da escala (p. ex. Peterson e Parker 1998, Lawton 1999, Pascual e Levin 1999, Petersen *et al.* 1999). Alguns dos avanços científicos mais importantes vêm de estudos em escalas diminutas e outros em escalas muito grandes. Entretanto, em um estudo integrado, especialmente um desenhado para subsidiar os políticos na tomada de decisões, haverá um leque limitado de escalas apropriadas para atender a necessidade geral do estudo, e estas escalas são, usualmente, grandes (Bradshaw 1998, Ormerod *et al.* 1999).

[Para decidir em que escala trabalhar, você deveria primeiro considerar a qual área/população/período de tempo/situações físicas/químicas/sociais você desejaria que os resultados se aplicassem.] Isto é chamado de seu universo de interesse. Potencialmente, todos os membros da equipe podem trabalhar em escalas similares, mas os cientistas tendem a copiar seus delineamentos amostrais de publicações recentes na literatura especializada de sua área. Pode ser difícil convencer um membro da equipe que ele deveria usar uma determinada escala de amostragem, se um famoso cientista acabou de publicar na revista de maior prestígio da área, um estudo a respeito do mesmo organismo (ou processo), usando uma escala de amostragem completamente diferente. Embora muito seja feito para inovar a ciência, a maioria dos cientistas acredita que serão criticados por seus colegas se desafiarem os dogmas estabelecidos, inclusive no que tange à escala de amostragem – e eles estão certos. Nos casos mais extremos, o líder da equipe deve decidir o quanto de prestígio pessoal ele está disposto a sacrificar pelo bem do projeto.

Para onde ir?

Em geral, os pesquisadores decidem qual método de análise irão usar copiando um disponível na literatura, e não consideram a especificidade de suas questões. Por causa disso, as análises tendem a definir as questões, ao invés de as questões definirem as análises (Yoccoz, N. G. 1991). Portanto, é importante que o líder e cada membro da equipe tenham algum conhecimento sobre os conceitos que embasam os tipos de análises mais comumente empregadas pela maioria dos cientistas. Afortunadamente, as análises estatísticas mais usadas são baseadas em poucos conceitos básicos. Para usá-los, precisaremos ser capazes de construir fluxogramas, interpretar gráficos simples e entender a filosofia que está por trás dos testes de hipóteses. Os fluxogramas que construímos mostram onde pensamos estar. Eles nos sugerem experimentos críticos ou observações que precisamos fazer. Platt (1964) disse que deveríamos "devotar de 30 a 60 minutos por dia para reflexão e análise, anotando explicitamente, em um caderno de notas permanente, as alternativas lógicas referentes aos nossos experimentos críticos". Infelizmente, estas habilidades não são ensinadas na maioria dos cursos de estatística. Tentaremos considerar alguns dos aspectos básicos destas habilidades nos próximos capítulos.

CAPÍTULO 3

As estatísticas geralmente escondem mais do que revelam padrões

Descrevendo a natureza: convenções "científicas" e algumas técnicas úteis

Neste capítulo vamos discutir algumas estatísticas simples, que são usadas para descrever dados ou as populações de medidas das quais eles foram obtidos. Técnicas estatísticas podem ajudar a elucidar padrões ocultos nos dados, se o modelo representado pelo nosso fluxograma for complexo, como veremos nos capítulos 8 e 10. Entretanto, a maioria das estatísticas encontradas na literatura científica relaciona-se com situações extremamente simples, e escondem mais do que revelam os padrões. Será útil para os leitores conhecer os termos, assim poderão ler a literatura científica e conversar com seus colegas. Mas, na verdade, apenas poucos conceitos serão importantes para os próximos capítulos, e estes são conceitos facilmente visualizados em gráficos bidimensionais.

Vamos considerar como um pesquisador poderia descrever seus dados sobre densidades de lagostins em riachos. O pesquisador amostrou 5 riachos e apresenta os dados como o número de lagostins por 100 m de riacho. Os dados poderiam ser apresentados como uma lista de densidades (1, 3, 4, 5, 7 lagostins/100m). Contudo, estes dados são freqüentemente apresentados juntamente com outros dados para comparação (veja capítulos 5 e 6). Portanto, vamos refletir sobre como eles podem ser apresentados da forma mais proveitosa. A forma mais simples seria colocá-los em um gráfico (figura 13).

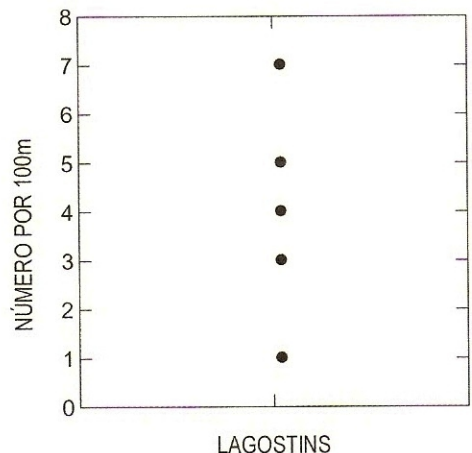
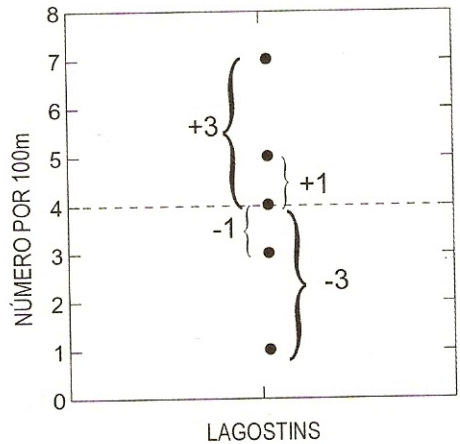
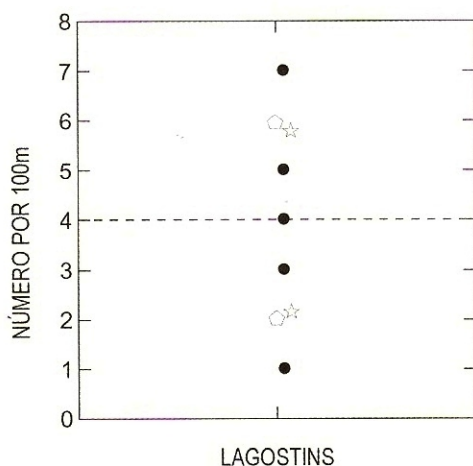
FIGURA 13

O gráfico ocupa muito espaço no papel, mas possibilita-nos avaliar instantaneamente o valor geral dos dados (média) e a variabilidade em torno da média. A maioria dos pesquisadores sabe instintivamente o que a média representa, mas não o quê várias estatísticas descritoras da variabilidade significam. A maioria dos iniciantes usaria a amplitude (a diferença entre o maior e o menor valor) para descrever a variabilidade, mas isto tem a desvantagem de usar apenas informação de dois dos cinco pontos. Uma alternativa um pouco mais complexa, mas ainda intuitiva, seria usar o desvio médio absoluto, ou seja, a média das distâncias de cada ponto até a média. A figura 14 mostra estas distâncias, exceto para a distância do valor "4", que está à distância zero da média. Poderíamos resumir os dados expressando a média = 4 e o desvio absoluto médio = 1,6. Usamos o desvio absoluto (módulo do desvio) porque a soma dos desvios simples será sempre zero, já que os desvios dos pontos situados abaixo da média sempre são negativos.

FIGURA 14

Com um pouco de prática, a maioria dos pesquisadores pode olhar para um gráfico e rapidamente estimar qual a região do gráfico compreendida entre a média menos um desvio absoluto e a média mais um desvio absoluto. O desvio absoluto é tão intuitivo que se poderia esperar que fosse a estatística mais usada para descrever a variabilidade dos dados. No entanto, ele é raramente usado, e uma outra estatística, o desvio padrão, ou alguma derivação dele, é o descritor de variabilidade mais freqüentemente empregado. O desvio padrão tem vantagens relacionadas com algumas análises mais complexas, que serão consideradas em capítulos posteriores. Entretanto, ele está longe de ser intuitivo. Entre uma a três vezes por ano, nos últimos 10 anos, temos solicitado, a nossos alunos de pós-graduação, que indiquem onde deve passar a linha que corresponde ao desvio padrão de cada lado da média em gráficos simples como o da figura 13. Poucos puderam aproximar a posição correta, e destes, quase nenhum soube explicar porque escolheram aquela posição. Muitos destes estudantes já haviam publicado trabalhos nos quais usaram o desvio padrão para descrever seus dados. É preocupante constatar que eles não sabiam o que estavam descrevendo e que não eram capazes de interpretar os resultados de outros autores que usaram o mesmo descritor.

O desvio padrão e as estatísticas relacionadas são usados tão freqüentemente que vale a pena despender um pouco de tempo para visualizar o que ele representa. Entretanto, lembrem-se que há inúmeros outros descritores de dispersão que podem ser mais apropriados em muitas situações (Iglewicz 1983). Assim como o desvio absoluto, o desvio padrão é baseado nas diferenças do valor de cada observação em relação à média, só que, no caso do desvio padrão, as

diferenças são elevadas ao quadrado. Se somarmos os desvios quadrados teremos uma quantidade chamada "soma de quadrados". Isto não é muito útil para descrever a variabilidade, porque, sendo uma soma, esta quantidade aumenta com cada observação extra. Contudo, se utilizarmos a média dos desvios quadrados, teremos um descritor que é independente do tamanho da amostra. Isto é chamado de variância. Mas, ao contrário da média absoluta, o desvio quadrado médio (variância) não é útil para descrever a variabilidade, porque está em uma potência diferente da dos dados originais e, provavelmente, nem irá caber em nossos gráficos. Por exemplo, é difícil de se visualizar lagostins com comprimento médio de cinco centímetros ± quatro centímetros quadrados ! Para trazer a medida de variação de volta à escala original em que as medidas foram tomadas, podemos extrair a raiz quadrada. A raiz quadrada da média dos desvios quadrados é chamada de desvio padrão. Isto parece uma rota sinuosa para se chegar a um valor que é diferente, mas não muito diferente, do desvio médio absoluto.

Os valores de um desvio médio absoluto (mostrados como estrelas na figura 15), e de um desvio padrão de cada lado da média (pentágonos) são quase idênticos. Ambos os desvios foram calculados para a população ao invés de para a amostra, mas não vamos explanar sobre as diferenças aqui, já que, para a interpretação gráfica a diferença é trivial, exceto para amostras extremamente pequenas. Vamos discutir sobre os usos do desvio padrão posteriormente, mas, por hora, basta interiorizar que o valor do desvio padrão usualmente não é muito diferente da média dos desvios absolutos. Portanto, podemos visualizar mentalmente a posição aproximada de um desvio padrão em cada lado da média, em um gráfico. Isto deverá ajudá-los a entender as descrições que outros pesquisadores fazem de seus dados.

FIGURA 15

Temos discutido a respeito de descritores de dados. Como os dados usualmente são amostras tiradas de populações muito maiores, estamos descrevendo apenas as amostras. Estes descritores, chamados "estatísticas" são freqüentemente usadas como estimativas dos descritores reais da populações estatísticas, estes últimos chamados de "parâmetros". Alguns autores usam terminologia um pouco diferente, mas, em geral, parâmetros são características das populações e estatísticas são estimativas dos parâmetros, baseadas em amostras. Note que, quando falamos "populações estatísticas" estamos nos referindo a populações de números, que podem ou não estar relacionadas com populações biológicas. A população de alturas de homens possui uma média muito maior do que a média da população de alturas de mulheres, e estamos

nos referindo a populações estatísticas. Mas quando se fala em "população humana", estamos nos referindo a uma população biológica que consiste aproximadamente em metade de homens e metade de mulheres. A altura média de uma espécie de antílope pode ser igual à altura média de homens, mas isto não significa que homens e antílopes pertençam à mesma população biológica.

O desvio padrão é um parâmetro útil para descrever a variabilidade em uma população de medidas, se a população de medidas tem uma distribuição de freqüências que conforma a distribuição teórica chamada de "normal". Neste caso, cerca de 68% da população está compreendida no intervalo de um desvio padrão em torno da média e cerca de 95% da população está compreendida no intervalo de dois desvios-padrões em torno da média. Se a distribuição não é normal, o desvio padrão não tem utilidade para descrever a variabilidade da população (Mosteller e Tukey 1968), um fato que parece ter sido esquecido pela maioria dos pesquisadores.

Podemos também calcular o desvio padrão de parâmetros, que é chamado de "erro padrão" (*EP*). Em geral, quando os autores tratam de erro padrão sem especificar de qual parâmetro, estão se referindo ao erro padrão da média. Para ilustrar isto, usamos um gerador de números aleatórios para produzir 300 medidas de uma distribuição com média e desvio padrão igual ao dos dados que aparecem na figura 15. Então, tiramos 60 amostras de cinco elementos, obtendo 60 médias, uma para cada amostra. A primeira amostra encontra-se ao lado das 60 médias na figura 16 e a média que corresponde a esta amostra está assinalada como um círculo cheio entre os círculos vazios que simbolizam as outras médias.

A amostra tem uma média (2,4) que é bastante diferente da média real das 300 medidas (4). Entretanto, a média das 60 médias (3,8) foi mais próxima da média real. Obviamente, a média de 60 amostras traz muito mais informação do que uma única amostra de cinco elementos. Com 60 médias, podemos também estimar o desvio padrão das médias (erro padrão) com razoável acurácia. O desvio padrão das 60 médias foi 0,77. Portanto, esperamos que cerca de 68% das médias tiradas desta população estejam compreendidas entre 4 + 0,77 e 4 - 0,77. A distribuição das médias tende à normalidade, mesmo se a distribuição das medidas originais não fosse normal. Isto é conseqüência do "Teorema do Limite Central" e é o que justifica o uso de testes estatísticos baseados na distribuição normal quando a população de medidas originais não segue esta distribuição.

FIGURA 16

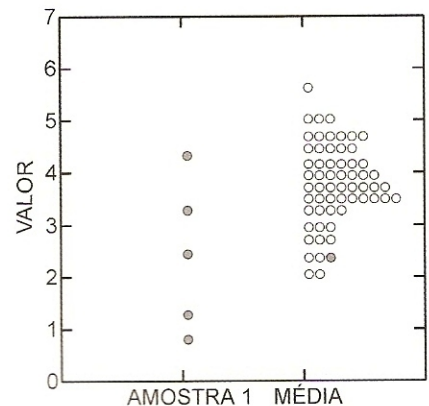
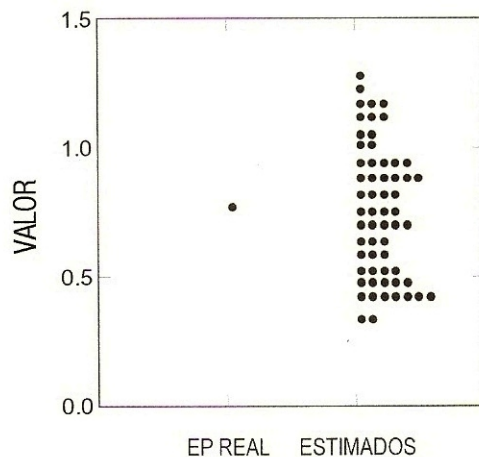
Se todas estas médias de médias e parâmetros de parâmetros os deixaram confusos, não se preocupem, não vamos usar muito estes conceitos, a não ser para examinar como alguns pesquisadores apresentam seus

dados, e como eles poderiam apresentá-los de forma mais simples e efetiva. Ninguém usa 60 amostras para calcular um erro padrão, porque os estatísticos nos dizem que podemos estimar o erro padrão a partir de uma única amostra. Há um erro padrão diferente para cada tamanho de amostra que desejarmos tirar. Entretanto, com uma única amostra de cinco elementos, podemos, teoricamente, estimar o desvio padrão das médias válido para um grande número de amostras de cinco elementos. A fórmula "mágica" é apenas dividir a estimativa do desvio padrão da população baseado na amostra pela raiz quadrada do número de observações da amostra. Em nosso caso, sabemos que o erro padrão das médias das amostras de cinco elementos está próximo de 0,77. A figura 17 apresenta a distribuição das 60 estimativas de erro padrão baseadas em nossas 60 amostras. Nós acrescentamos nesta figura o erro padrão baseado no desvio padrão das 60 médias chamando-o de "real", porque ele deve ser muito próximo do valor real do erro padrão, que é desconhecido.

FIGURA 17

Obviamente, as estimativas dos erros padrões baseadas em amostras de cinco elementos são imprecisas. Muitas estão distantes da melhor estimativa. Quanto maior o tamanho da amostra, mais acuradas as estimativas obtidas pela fórmula mágica. Entretanto, raramente elas podem ser melhor interpretadas por um estatístico do que um gráfico com os dados brutos, e a maioria dos pesquisadores (os leitores de seus trabalhos) têm apenas uma vaga idéia do que eles representam. É possível inferir intervalos de confiança baseados na estimativa do erro padrão e no tamanho da amostra. Os estudantes interpretam um intervalo de confiança de 95% (IC 95) em torno da média como se houvesse 95% de chance de que o parâmetro real estivesse neste intervalo. Do ponto de vista da estatística clássica, esta explicação não tem sentido. Ou a média cai no intervalo ou não. A interpretação correta do intervalo de confiança de Neyman (Neyman 1937) é que, se um experimento for repetido muitas vezes, com o mesmo tamanho amostral, e o intervalo de confiança for calculado para cada um deles, aproximadamente 95% dos intervalos calculados irão incluir a média real (Bard 1974). Entretanto, estes intervalos de confiança só funcionam se forem aplicados em todos os casos. Se intervalos de confiança forem aplicados apenas após algum teste estatístico ter tido resultado significativo, como é a prática corrente, então os intervalos de confiança apresentados nos livros de texto são pequenos demais (Meeks e D'Agostino 1983). Poucos pesquisadores entendem estes conceitos. De qualquer modo, a maioria dos erros padrões é apresentada em gráficos sem qualquer informação do tamanho da amostra na legenda (Magnusson, 2000a). Freqüentemente, há diferentes tamanhos de amostras em diferentes partes do gráfico. Isto demonstra cultura "científica", mas não comunica muito mais do que isto.

A seguir, ilustraremos diferentes métodos de apresentar dados em gráficos e deixar que os leitores decidam quais os melhores para comunicar informações sobre os dados obtidos. Primeiro, vamos considerar um gráfico mostrando dados com uma amostra pequena (figura 18).

A figura 18a mostra os dados de densidades de lagostins em diferentes riachos. A figura 18b foi baseada nos mesmos dados da figura 18a. Ela mostra algumas estatísticas usadas para sumariar dados e que freqüentemente aparecem em gráficos. O desvio padrão *(DP)* foi bem diferente do erro padrão (*EP*), mas não muito do intervalo de confiança de 95% (IC-95%). Se o leitor não estiver seguro de que ele mesmo e seus leitores entenderão as implicações das três estatísticas sumárias, então deveria considerar apresentar os dados da mesma forma que os apresentados na figura 18a. Gráficos como o da figura 18a são chamados de "gráficos de dispersão" e podem apresentar uma quantidade surpreendente de informação. A figura 19a mostra um gráfico de dispersão do número de lagostins onde aparecem os dados de um pesquisador que amostrou 40 riachos com peixes e 40 riachos sem peixes. Este gráfico é um pouco difícil de interpretar, porque alguns pontos podem se sobrepor e obscurecer outros. Na figura 19b separamos os pontos. Este último gráfico é tecnicamente um histograma de pontos, mas em termos de comunicação pode ser considerado um gráfico de dispersão com os pontos separados de modo que nenhum ponto obscureça os demais.

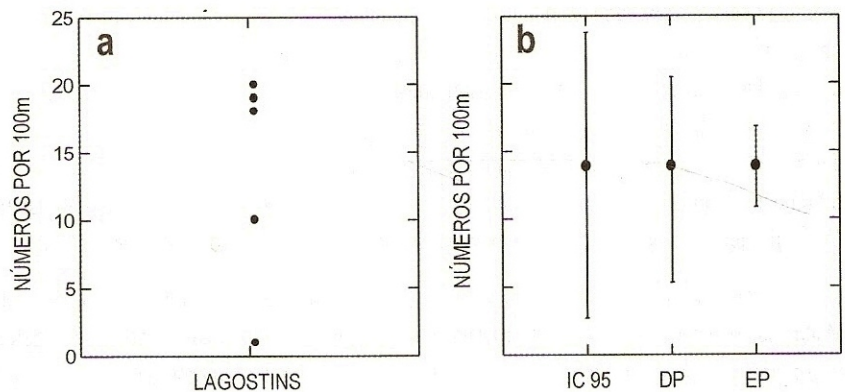



FIGURA 18

Note que este gráfico mostra quase todos os detalhes a respeito dos dados, mas muitos cientistas não os consideram "científicos". Afinal, qualquer pessoa, até os não-cientistas podem avaliá-los. Seus colegas em uma equipe de pesquisa provavelmente irão preferir gráficos que demonstrem sua cultura científica, mesmo que estes gráficos distorçam a informação. A figura 20a ilustra um gráfico de barras representando os mesmos dados da figura 19, incluindo barras de erros para os erros padrões. Este tipo de gráfico efetivamente esconde toda a informação a respeito da quantidade de dados que coletamos. Os estatísticos dirão que não há problemas para se interpretar as barras de erros, porque suas premissas a respeito do mundo real estão sempre corretas e qualquer biólogo competente tem um

sentimento intuitivo do que um erro padrão representa e deve ser capaz de reconstituir o gráfico original em sua cabeça. Contudo, temos ministrado cursos de estatística para estudantes de graduação e pós-graduação e temos sido consultados por pesquisadores experientes durante um período de mais de duas décadas, e podemos assegurá-los de que nenhuma destas duas afirmações é verdadeira. Se você deseja esconder seus dados, coloque-os em gráficos de barras. O "box plot" da figura 20b é um pouco melhor. Entretanto, poucos pesquisadores sabem como interpretá-los e ainda perdemos a informação a respeito de quantos dados foram utilizados para produzir o gráfico. "Box plots" são uma boa alternativa quando você tem tantos dados que não podem ser apresentados em um gráfico de dispersão (Tukey 1972), mas ao contrário deste, não têm utilidade para se planejar a amostragem (capítulo 4). Use-os como última alternativa, nunca como o método padrão de apresentar os resultados, sejam dados hipotéticos ou não.

FIGURA 19

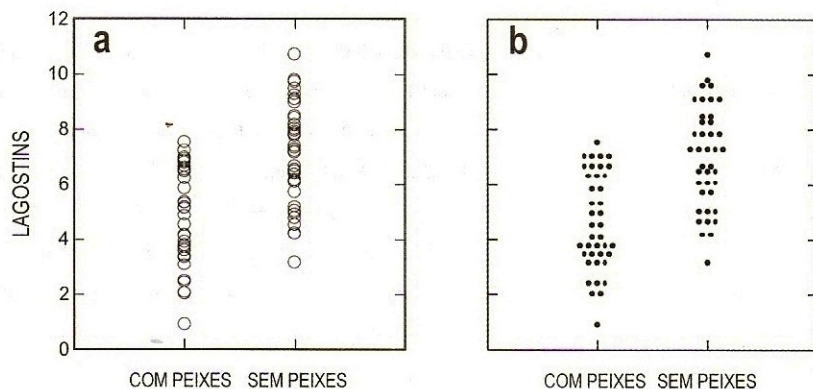
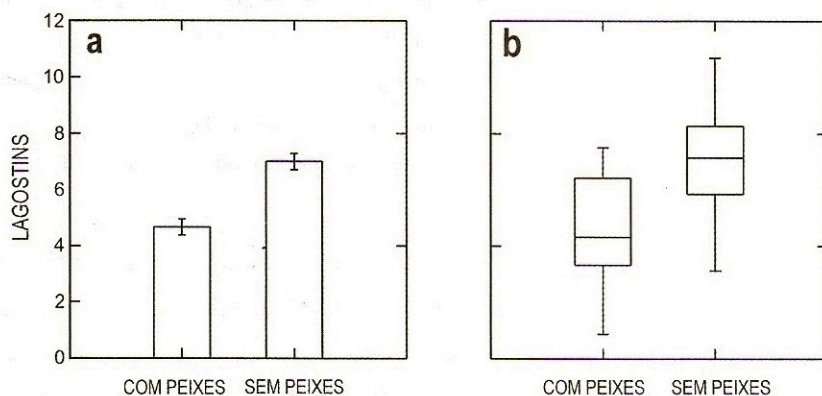



FIGURA 20

Repetições geradas no computador

É importante lembrar que algumas medidas de variabilidade geradas por computadores podem não ser aquelas nas quais você está interessado. Há métodos para o cálculo de erros padrões baseados em transectos de linha (p. ex. Krebs 1998). Entretanto, lembre-se que os erros padrões baseados em métodos de distância referem-se aos erros padrões *daquele* transecto. Ele estima a variabilidade esperada se você repetir o mesmo transecto, mas não informa nada a respeito do erro padrão esperado caso você replicasse os transectos na área de interesse. Isto precisa ser estimado por outros métodos (Caughley e Sinclair 1994). Primatólogos freqüentemente apresentam erros padrões de densidades de macacos em reservas, baseados em um ou dois transectos. Isto não faz sentido. Eles confundem a variabilidade da amostragem repetida daquele transecto com a variabilidade entre transectos. O método poderia ser aproximadamente válido se o transecto fosse tão comprido que amostrasse quase toda a reserva. Entretanto, este seria um delineamento amostral muito ineficiente. Quando apresentamos uma medida de variabilidade, precisamos estar certos de que ela se refere ao nosso universo de interesse.

Um problema relacionado é a estimativa do erro padrão associado à estimativa do tamanho da população em estudos de marcação e recaptura (Abuabara e Petrere 1997, Krebs 1998). Métodos de marca-recaptura estimam o número de indivíduos na população que são suscetíveis de serem capturados. Converter isto em uma estimativa de densidade é complicado, a não ser que a população ocupe uma área restrita e todos os animais na população sejam suscetíveis de serem capturados (Anderson *et al.* 1983). Animais freqüentemente variam sua área de vida sazonalmente ou em períodos interanuais, devido à variação nos recursos. A estimativa do tamanho da população e seus erros padrões podem estar estatisticamente corretas, mas a maioria dos ecólogos está interessada na estimativa da densidade, não do tamanho da população que pode ser capturada em armadilhas.

Neste capítulo, estivemos preocupados com as estatísticas usadas para resumir dados. Vamos usar alguns dos termos discutidos aqui nos capítulos posteriores, onde a variação será freqüentemente expressa como soma de quadrados, ou variância. Entretanto, o leitor não precisa ser capaz de visualizar a soma de quadrados ou a variância para usá-las. As outras estatísticas que sumarizam dados raramente são úteis, exceto para mostrar cultura acadêmica, e não deveriam ser usadas, a não ser quando tabelas ou gráficos de dispersão não sejam alternativas viáveis. Não se preocupe se não puder visualizá-las. Elas são como as roupas do rei, somente os verdadeiramente honestos vão admitir que não podem vê-las.

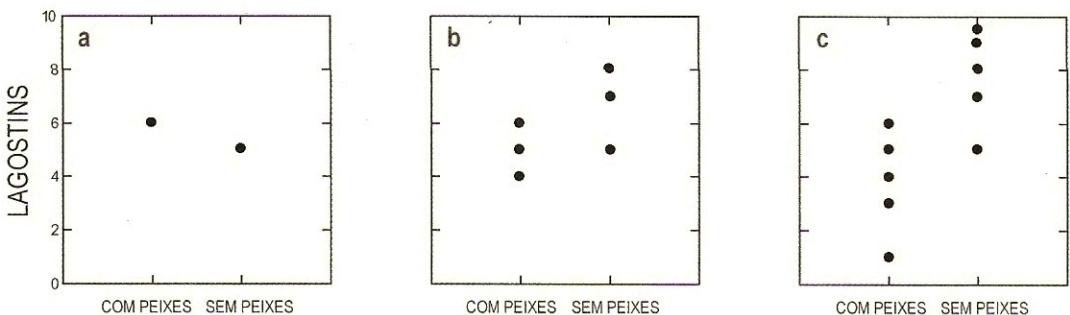
CAPÍTULO 4

[Decidir quanta informação é adequado, é também a arte do cientista.]

Quanta evidência é necessária?

Nosso artista pintando a paisagem não usa uma tela maior que a parede da sala. Ele também não tentará contar a estória de 10 maneiras diferentes na mesma tela. Se ele deseja se comunicar e não morrer de fome antes de completar seu trabalho, terá que decidir o quanto é suficiente. Decidir o quanto de informação é adequado, é também a arte do cientista. Vamos começar considerando uma questão simples e um gráfico simples.

Nosso cientista, estudando o lagostim, decide testar se locais sem peixes predadores têm mais lagostins do que locais com peixes predadores. Ele conta o número de lagostins em seções de riachos com e sem peixes predadores, mas está inseguro sobre quantas seções deve amostrar. Obviamente, uma seção de cada tipo não vai nos dizer muito (fig. 21a). Então ele aumenta o número para três de cada tipo, mas ainda permanece a dúvida (fig.21b). Entretanto, com cinco seções de riachos de cada categoria, já não resta muita dúvida (fig. 21c).

FIGURA 21

Este processo, que Dytham (1999) chamou de "coleta de dados substitutos" (*"collecting dummy data"*) parece ser trivial, mas é uma das maneiras mais poderosas de se planejar a pesquisa. Connolly *et al.* (2001) chamam esse processo de experimento de pensamento *(thought experiment)*. Há muitas fórmulas matemáticas elegantes para decidir quantas observações são necessárias para se detectar um efeito de uma dada magnitude (p. ex. Krebs 1989), entretanto todas requerem amostras preliminares e a maioria somente pode ser aplicada em situações trivialmente simples. Em geral, solicitar a um pesquisador experiente que esboce um gráfico mostrando a variabilidade esperada para aqueles dados é quase tão bom quanto qualquer um dos métodos que empregam computações matemáticas. Apenas vá aumentando o número de pontos no gráfico, até que o padrão pareça convincente.

Há algumas "dicas" que podem ser usadas, se você não conseguir extrair dos seus colegas de equipe informação útil em relação às questões na escala proposta para o projeto. Dados ecológicos usualmente mostram variabilidade semelhante àquela da figura 21. Assim, para comparações entre categorias, deveria haver pelo menos quatro observações por categoria ou preferivelmente mais. Entretanto, geralmente não há vantagem em ter mais que dez observações por categoria, a não ser que os dados possam ser coletados facilmente e a um baixo custo, ou há alguma razão especial para se desejar detectar diferenças muito pequenas entre as categorias. Vamos oferecer outras dicas de como decidir o tamanho da amostra nos próximos capítulos, mas quatro observações por categoria vão funcionar bem para a maioria dos casos.

Qual a qualidade da informação?

Para responder a questão "Quanta informação é suficiente?" precisamos primeiro responder "Qual a qualidade da informação?". Esta é uma pergunta importante, porque as questões estatísticas devem refletir as questões científicas. No início da década de 80, o estatístico Stuart Hurlbert alertou o mundo a respeito de um tipo de erro que vinha ocorrendo em grande parte das análises ecológicas, e que ele denominou de "pseudo-repetição". Repetições são o que os estatísticos chamam de observações independentes. Geralmente, espera-se que a quantidade de informação disponível aumente com o número de observações, mas nem sempre é assim. Quando uma nova observação fornece apenas a mesma informação que tínhamos em observações anteriores, ela não aumentou a quantidade total de conhecimento disponível para nós, e pode nos confundir e nos fazer acreditar que dispomos de mais informação do que realmente temos. Porque esta observação não é uma repetição real, no sentido de fornecer mais informação, Hurlbert denominou-a "pseudo-repetição", que significa falsa repetição.

Hurlbert (1984) forneceu muitos exemplos bons, mas vamos continuar com nosso exemplo do lagostim. Lembrem que a questão biológica era determinar se a presença de peixes predadores influencia na densidade dos lagostins. O que aconteceria se nosso biólogo amostrasse cinco seções de um riacho sem peixes e cinco seções de um riacho com peixes

predadores? Todas as seções do primeiro riacho poderiam ter menos lagostins porque este riacho era menos produtivo, ou porque fora poluído ou por causa de uma epidemia no passado recente ou qualquer um de uma multidão de fatores possíveis de afetarem lagostins. Para a nossa questão, uma observação na densidade de lagostins em um riacho obviamente não é independente de outras observações no mesmo riacho. Portanto, as cinco observações não carregam mais informações do que uma única observação.

Este erro simples permeia a literatura científica. Kruskal (1988) fornece vários exemplos de falta de independência em situações que não têm nada a ver com ciência. Cada coisa está conectada com tudo o mais e é extremamente difícil determinar quando as observações são realmente independentes. A engenhosidade dos cientistas revela-se em sua capacidade de coletar observações realmente independentes em relação a suas questões, mas esta é uma qualidade rara. Portanto, os membros de grupos integrados precisam se certificar que as observações que estão sendo feitas por seus colegas de equipe não são pseudo-repetições em relação à questão global. Isto é uma tarefa difícil, porque **nenhuma observação é inerentemente válida ou inválida.** Uma pseudo-repetição para uma questão pode ser uma repetição válida para outra. Por exemplo, se o nosso biólogo estivesse interessado em diferenças nas densidades do lagostim entre os dois riachos (e apenas entre os dois), as 5 observações em cada corpo d'água poderiam ser repetições perfeitamente válidas, cada uma trazendo mais informação a respeito da densidade do lagostim em cada riacho.

As pseudo-repetições podem ser espaciais, temporais, filogenéticas ou técnicas. Nós discutimos um exemplo de pseudo-repetição espacial e Hurlbert (1984) fornece inúmeros outros. Este é o tipo de pseudo-repetição mais comum, mas também é a que pode ser evitada com maior facilidade. Os membros da equipe vão atribuir suas pseudo-repetições espaciais a dificuldades logísticas, mas o tempo e dinheiro gastos em transporte usualmente são mais do que compensados pela solidez das conclusões que podem ser tiradas de evidências de alta qualidade, coletadas em uma ampla escala espacial.

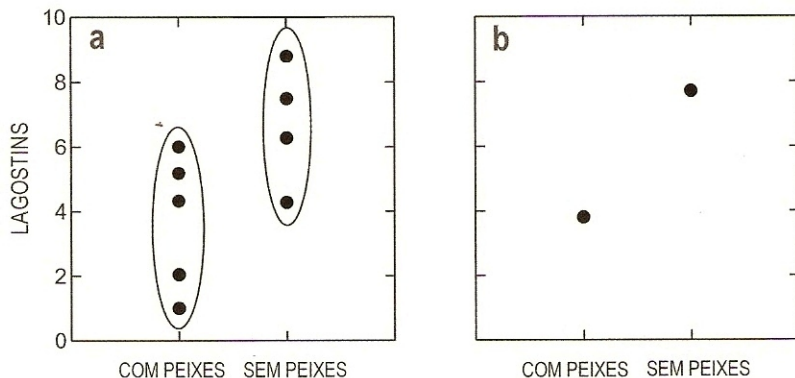
Pseudo-repetições temporais são mais difíceis de se detectar e evitar. Ocorrem porque o estado de um sistema não pode mudar instantaneamente. Se uma árvore está produzindo frutos em um mês chuvoso, provavelmente ainda estará produzindo frutos no mês seguinte, e a precipitação nesse mês provavelmente ainda será alta. Parece haver uma grande associação entre alta precipitação e produção mensal de frutos, embora o evento que determinou o início do processo de frutificação tenha ocorrido muitos meses antes, ainda durante a estação seca.

Por esta razão, as causas de sazonalidade não podem ser demonstradas em um único lugar em apenas um ciclo. Décadas de trabalho em um único lugar ou estudos em muitos lugares com diferentes padrões climáticos são necessários para se determinar as causas da sazonalidade. A despeito disso, muitos membros da equipe vão querer estudar sazonalidade, porque eles acreditam que os efeitos da sazonalidade são tão óbvios que eles certamente conseguirão resultados "significativos". As causas de variações temporais são extremamente difíceis de estudar (p. ex. Powell & Steele 1995, von Ende 1993) e os membros de equipes deveriam procurar o aconselhamento de estatísticos especializados, antes de incluírem variação temporal nos objetivos gerais de seus estudos.

Pseudo-repetição filogenética é um tópico complexo e, na maioria das vezes, está associada a estudos onde espécies são as unidades amostrais. Os leitores mais interessados podem consultar Garland *et al.* (1992), um bom ponto de partida para este tópico. Rohlf (2001) compara a utilidade e eficiência dos vários métodos de análise disponiveis. Estudos com sementes, girinos e larvas de insetos freqüentemente sofrem de pseudo-repetição filogenética, porque os organismos usados nos experimentos têm estreito parentesco. Sementes de uma única árvore, girinos de uma mesma desova ou larvas de insetos encontradas em um local em particular são freqüentemente muito similares, por causa do material genético ou efeito do aprovisionamento materno e usualmente não apresentam todo o espectro de respostas exibido por populações maiores. Em vez de tentar fazer correções para excluir efeitos filogenéticos, usualmente é preferível tentar coletar indivíduos não relacionados entre si para os experimentos. Se a escala de amostragem for apropriada para a questão, isto deverá acontecer quase naturalmente.

Pseudo-repetição técnica ocorre quando diferentes observadores ou instrumentos são usados em diferentes partes do experimento. Quando não é detectada, ela recai no que Hurlbert chama de "intrusão demoníaca", mas há poucas desculpas para este tipo de erro na maioria dos estudos bem planejados. Ocasionalmente, há falhas de equipamento ou mudança da pessoa que está coletando os dados que podem alterar os resultados. Nestes casos, admitir honestamente a possibilidade de pseudo-repetição é a única opção.

Retornando ao nosso gráfico, o efeito de pseudo-repetição é reduzir o número de observações realmente independentes, que podem ser representadas por pontos no gráfico. Se apenas dois riachos foram amostrados, todos os pontos mostrados na figura 22a ficam reduzidos aos dois pontos da figura 22b. Poucos biólogos, menos estatísticos, e nenhum computador, serão capazes de perceber que a maior parte da informação mostrada na figura 22a não pode ser interpretada em relação à questão inicial. Dados somente são úteis se fornecem informação. Se nossos gráficos são enganadores, nossas análises também o serão e não teremos feito uma apreciação honesta das evidências. Há algumas técnicas estatísticas que podemos usar para levar em consideração o fato de que nem todas as observações foram independentes (p. ex. análise de variância hierárquica), mas somente serão úteis se os pesquisadores puderem reconhecer a falta de independência e informar este fato ao estatístico ou ao computador. Nenhuma destas técnicas é tão boa quanto reconhecer que o delineamento proposto leva a pseudo-repetições e re-delinear o sistema de amostragem para evitá-las.

FIGURA 22

Para decidir quanta evidência é suficiente, temos de pensar em quantos pontos teremos que colocar em cada gráfico e nos assegurarmos de que estes pontos carregam informação independente em relação à nossa questão. Não é preciso ser um estatístico ou um biólogo para fazer isto, e o processo é notadamente simples. Entretanto, pela falta deste primeiro passo, a maioria dos resultados publicados em trabalhos científicos é, na melhor das hipóteses, não interpretável e na pior, enganosa. Vamos dar um passo adiante e fazer considerações sobre um teste do qual quase todos já ouviram falar, o teste do qui-quadrado em tabelas de contingência. Análises de tabelas de contingência quase nunca são apropriadas para questões estatísticas (Hulbert 1984, Magnusson 1999), e isto se relaciona com o que é uma observação independente em relação a uma questão. Um biólogo estudando a dieta de uma espécie de peixe formula a pergunta se a dieta dos adultos é diferente da dos alevinos, e produz a tabela 2, que contrasta o número de copépodos e algas encontrados nos estômagos de alevinos e peixes adultos.

O biólogo sabe que será praticamente impossível publicar os resultados sem um teste estatístico e então aplica um teste de tabela de contingência, que resulta nos seguintes hieróglifos: $\chi^2_1 = 6030$, $P << 0.001$. Não se preocupe se você não sabe o que eles significam. A maioria dos leitores vai rever o que eles achavam que isto significava, após lerem mais alguns capítulos. O importante aqui é que a maioria dos ecólogos iria interpretar este resultado como indicativo de que é muito improvável que peixes adultos e alevinos tenham a mesma dieta. Contudo, esta impressão muda quando vemos o caderno de anotações do biólogo (tabela 3).

Agora, ninguém acredita em uma diferença geral na dieta de adultos e alevinos. Por acaso, um adulto encontrou um grupo de copépodos e um alevino fez um banquete em um aglomerado de algas. Por quê o teste nos deu uma resposta falsa? Porque a análise assumiu que o registro de cada copépodo ou alga foi independente dos outros. Os dados não refletem contingência. Nós poderíamos levar a cabo um teste de contingência válido, se selecionássemos ao acaso apenas um item por cada estômago de peixe, e descartássemos todo o resto da amostra. Isto iria requerer 6194 indivíduos de peixes, cada um coletado em um cardume diferente. Obviamente, esta seria uma maneira muito ineficiente de atacar a questão. É justamente o fato de as tabelas de contingência requererem observações independentes e somente acumularem informação na forma de presença/ausência que as fazem tão ineficientes para a maioria das questões ecológicas.

TABELA 2

	COPÉPODOS	ALGAS
ADULTOS	3217	18
ALEVINOS	23	2936

TABELA 3

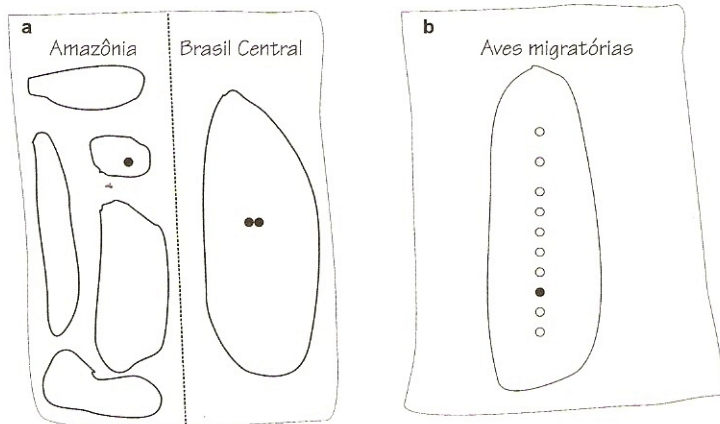
	COPÉPODOS	ALGAS
ADULTO 1	0	6
ADULTO 2	3211	7
ADULTO 3	6	5
ALEVINO 1	8	2906
ALEVINO 2	8	1
ALEVINO 3	7	29

Quando o leitor encontrar uma análise de tabela de contingência, pode suspeitar que o pesquisador cometeu pseudo-repetição e que as inferências estatísticas são desprovidas de sentido. Há uma pequena chance de a análise estar correta e neste caso o delineamento amostral provavelmente foi muito ineficiente. Agora o leitor pode ver porque cursos de estatística básica freqüentemente examinam apenas probabilidades referentes a se retirar bolas pretas e brancas de um barril. Barris não se reproduzem, não nadam em cardumes e não mudam muito com o tempo. Por isso eles são fáceis de ser modelados matematicamente. Entretanto, eles podem não ter muita relevância para a maioria dos ecólogos.

A melhor maneira de evitar pseudo-repetições é desenhar um mapa conceitual da distribuição dos objetos de interesse em seu estudo. Cada ponto no seu diagrama deve representar uma unidade de amostragem em potencial. A unidade de amostragem pode ter diferentes nomes, dependendo da disciplina. Pode ser chamada de transecto, parcela, grupo focal, grão, pixel, etc.. Unidades de amostragem são freqüentemente relacionadas com a área, mas pode ser uma espécie, um intervalo de tempo, unidades experimentais ou outros objetos. Elas são sempre a menor unidade que se pode medir, que ainda faz sentido em relação à questão investigada.

O diagrama deve cobrir todo o universo de interesse. Este pode ser chamado de universo de inferência, escala de inferência, escopo, extensão, imagem e outros nomes. Estes diagramas não precisam ser acurados ou artísticos; servem apenas para proporcionar uma impressão geral.

Dois exemplos reais podem ilustrar a técnica. Uma estudante quer comparar o comportamento de pássaros nos cerrados amazônicos e no Brasil central. Ela estuda muitos indivíduos, mas quando mapeia os cerrados de interesse (a figura 23a é uma representação conceitual do mapa), ela descobre que seu universo amostral não é o mesmo que o universo de interesse. Note que este mapa conceitual simples, com formas ovais representando áreas de cerrado, uma linha reta

FIGURA 23

representando a borda entre os dois biomas e três pontos pretos representando áreas amostrais foi suficiente para mostrar que o universo de amostragem não corresponde ao universo de interesse da questão original.

A figura 23b mostra o universo de interesse de um estudante que trabalha com pássaros migratórios e deseja fazer inferências sobre migrações de longas distâncias. Embora ele se interesse por várias espécies de pássaros (representadas pelas bolas cinzas), amostrou apenas uma das espécies que fazem aquela migração (representada pela bolinha preta).

Um diagrama simples como o da figura 23b, com espécies de interesse representadas por pontos, foi suficiente para mostrar o quanto o universo de amostragem do estudante era limitado em relação ao seu universo de interesse. Conseqüentemente, ele compreendeu que, antes de fazer inferências fortes ele deveria restringir sua questão a apenas uma espécie ou aumentar o número de espécies amostradas. Mapas conceituais quase sempre irão dar uma boa idéia de se o pesquisador está usando os dados apenas como uma vaga pista (sem repetições verdadeiras), como uma inferência forte (inúmeras evidências independentes), ou algum meio termo. O que parece ser mapear as unidades amostrais, na verdade é um processo para se decidir qual a questão que será atacada.

Nos capítulos seguintes, vamos discutir como **grande parte das análises estatísticas pode ser vista como métodos de se reduzir problemas complexos a problemas com duas dimensões, de forma que possam ser apresentados em simples gráficos de dispersão.** Se os membros de equipes de pesquisa não puderem representar os resultados esperados de suas análises em forma de gráficos simples, eles não compreendem aquelas análises e não deveriam usá-las. Entretanto, quando confrontados com sua incapacidade de produzir gráficos bidimensionais que representem seus resultados, os líderes de equipe irão clamar que o objetivo de um estudo científico não é produzir gráficos simples, que podem ser entendidos por qualquer pessoa. Eles dirão que o objetivo dos estudos científicos é estimar probabilidades. Nos próximos capítulos vamos considerar a estranha definição de probabilidade usada pela maioria dos cientistas.

CAPÍTULO 5

> [O fundamento da teoria Popperiana é que não se pode provar nada, apenas "desprovar"]

Quando *improvável* significa *bem possível*

Os estatísticos vêm usando, por muito tempo, uma definição *sui generis* de probabilidade, que vai contra a nossa intuição (Platt 1964). Entretanto, a aceitação geral desta definição remonta a Sir Karl Popper, nas primeiras décadas do século XX. Popper era um austríaco, e por isso foi surpreendente que ele tenha sido agraciado com o título de "Cavalheiro" pela Rainha da Inglaterra. Sua filosofia teve larga aplicação nos ramos da política, ciências sociais e estudo do aprendizado. Contudo, muitos cientistas não se dão conta disto e usam a filosofia Popperiana apenas como um árbitro imparcial para determinar um mundo "objetivo".

O fundamento da teoria Popperiana é que não se pode provar nada, apenas "desprovar". Para realmente entender isto, é preciso ler muito mais do que poderíamos apresentar aqui. Aqueles interessados em conhecer mais a respeito de Popper e suas idéias, em uma prosa colorida, embora prolíxa, devem ler a sua "Unended Quest: an Intellectual Autobiography." (Popper 1976). Aqueles que não tiverem tanto tempo, podem consultar "Popper" de Magee (1982).

No campo político, Popper desmantelou a justificativa "científica" nazista para o genocídio de judeus, e por isso foi agraciado por Sua Majestade. Suas idéias sobre o processo de aprendizado humano abriram, de muitas maneiras, caminhos para a moderna sociobiologia. A filosofia Popperiana é a base para a abordagem de chaves dicotômicas de decisão em estudos de planejamento, que tem sido reconhecida como uma das abordagens mais poderosas na ciência (Platt 1964). Sua filosofia é o eixo central de quase toda a literatura moderna sobre estatística, e é este o aspecto que vamos considerar neste volume. Entretanto, a filosofia Popperiana, e especialmente a abordagem mais restrita chamada "freqüentista", não é o único ou necessariamente o melhor método científico. Pickett e colaboradores (1994) fornecem uma boa introdução sobre outras maneiras de se olhar o mundo.

Hilborn e Mangel (1997) escreveram um grande livro chamado "*The Ecological Detective*", no qual ofereceram uma visão geral de diferentes meios de se abordar fenômenos ecológicos e contrastaram algumas abordagens de investigação. Ao invés de considerarmos diferentes agendas de pesquisa como abordagens diferentes, podemos interpretá-las apenas como diferenças nas ênfases. Embora Hilborn e Mangel (1997) foram avessos a chamar seus modelos de "hipóteses" e ministraram o paradigma Bayesiano, eles iniciaram sua discussão com a premissa de que os leitores já conheciam tudo de que tratamos neste livro. Recomendamos que os professores de estatística leiam o livro de Hilborn e Mangel antes de ministrarem seus cursos e que os alunos leiam-no após a completa leitura deste livro.

Popper ensina que só aprendemos quando erramos. Considere uma meninha recém-nascida. Ela está programada para esperar alguma coisa do mundo. Se ela não tivesse nenhum programa em seu cérebro, seria como um computador sem um sistema operacional – totalmente morto. Vamos imaginar que a criança é programada para acreditar que o mundo é macio (isto é hipotético, já que os autores admitem que não têm lembranças tão antigas). Os pais da criança fazem o possível para que tudo o que toque a menina seja cor-de-rosa e macio em seus primeiros dias de vida e, portanto, ela não tem razões para rejeitar sua preconcepção de que o mundo todo é macio. Entretanto, note que não importa quantos objetos macios ela tenha tocado, isto nunca provará que o mundo é macio. Basta tocar um único objeto rígido, talvez um brinquedo de cor azul, para rejeitar toda sua conjectura sobre um mundo macio. A criança descobre que estava errada e aprende. Ela criará uma nova conjectura, talvez que apenas objetos azuis sejam rígidos, e somente irá adiante quando houver alguma razão para refutar esta nova conjectura.

Este exemplo ilustra o fato de que é fácil obter evidência para rejeitar alguma coisa. Freqüentemente uma única observação, que contrarie as espectativas, será suficiente. Entretanto, nenhuma quantidade de corroboração provará que aquela conjectura seja correta. Durante milhares de anos, todas as observações corroboravam a afirmativa de Aristótoles de que um corpo permanece parado na ausência de forças, até que Newton finalmente a tenha refutado. Segundo Newton, a aceleração, e não o movimento, demandava forças, e a aceleração de um corpo era diretamente dependente da força exercida sobre ele, e inversamente dependente de sua massa. Bastou alguns séculos para que Eistein refutasse Newton, demonstrando que essa relação não é válida para todos os casos (*i.e.* para escalas muito grandes ou muito pequenas). Hoje, os físicos se esforçam para refutar Einstein e empurrar a ciência adiante.

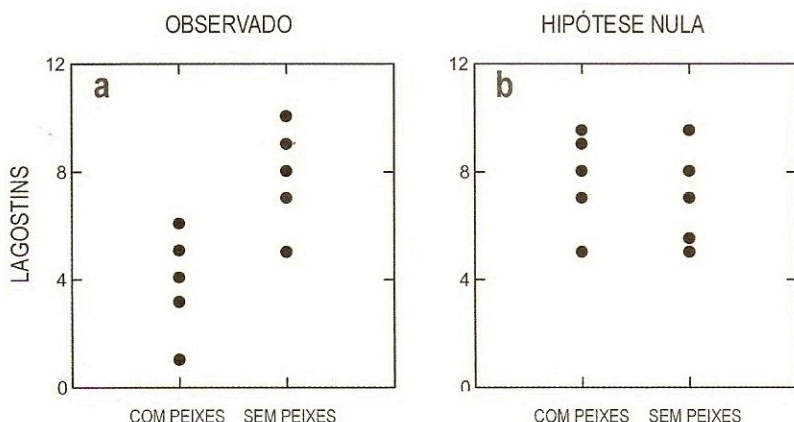
A estatística Popperiana se baseia nesta linha de raciocínio, de que é mais fácil refutar do que provar alguma assertiva, embora o embasamento da maioria dos livros convencionais de estatística seja muito mais restritivo do que a teoria geral de Popper, uma linha que tem sido chamada de "freqüentista". Os adeptos da estatística convencional (freqüentista) não perguntam qual é a probabilidade de estarem certos, mas a probabilidade de estarem errados. Para fazer isto, eles começam estabelecendo uma hipótese nula. **Uma hipótese nula é uma assertiva de como o mundo deveria ser, se nossa suposição estivesse errada.** Fizemos a conjectura de que a presença de peixes afetava a densidade de lagostim e o nosso gráfico foi consistente com nossa conjectura. Nossa hipótese nula era que não havia diferenças entre as densidades de lagostins entre os riachos com e sem peixes. A figura 24a ilustra nosso gráfico inicial, construído sob a hipótese de que os peixes afetam as densidades

do lagostim. A figura 24b ilustra como imaginamos que o gráfico deveria parecer se a hipótese nula (não há diferença nas densidades de lagostins em riachos com e sem peixes) fosse "correta".

A arte do cientista é ser capaz de vizualizar gráficos que representem a hipótese nula e comparar esta idealização com o gráfico obtido com dados reais. Entretanto, há muitos gráficos que podem representar a hipótese nula. Poderíamos ter amostrado outros 10 riachos e o resultado seria ligeiramente diferente. A estatística inferencial diz respeito a lidar com muitas possibilidades diferentes sob a hipótese nula.

Obviamente não podemos ajustar muitos gráficos em uma única página, portanto, iremos olhar para apenas uma coisa em cada gráfico para fazer a comparação. Esta medida de cada gráfico é chamada de uma estatística. Agora, precisaremos fazer algumas operações, mas serão contas simples. Para os dados mostrados na figura 24a, calculamos a média das densidades nos riachos com peixes (3,8) e nos riachos sem peixes (7,7), e a diferença entre as médias, que chamaremos de "*DIF*" foi igual a -3,9. Para os dados hipotéticos mostrados na figura 24b, a diferença entre a média da densidade de lagostins nos riachos com peixes (7,7) e sem peixes (7,0) foi igual a 0,7 (*i.e. DIF* = 0,7).

É intuitivo: quando peixes afetam as densidades de lagostins, esperamos maiores diferenças absolutas entre as médias das densidades em riachos com e sem peixes do que quando os peixes não afetam as densidades de lagostins. Entretanto, um cético poderia dizer que a associação entre a densidade de lagostins e a presença de peixes foi acidental. Ao amostrarmos apenas cinco riachos de cada tipo, poderíamos ter selecionado inadvertidamente cinco riachos com peixes que, por acaso, tinham menos lagostins do que os riachos sem peixes. Devemos reconhecer que o crítico tem sua razão. Se apenas um riacho não tivesse lagostins, e por puro acaso fosse justamente um dos riachos da categoria dos "com peixes", poderia ser suficiente para abaixar a média geral das densidades de lagostins em riachos com peixes.

FIGURA 24

Para avaliar este argumento, vamos criar uma hipótese mais específica a respeito de nossas observações. Poderemos dizer que esperamos uma diferença nas densidades dos lagostins, entre riachos com e sem peixes, maior que a esperada para uma associação ao acaso entre a presença ou ausência de peixes. A hipótese nula estabelece que ainda esperamos alguma diferença entre as médias das densidades de lagostins nas duas categorias de riachos, mas que esta diferença não é maior que a esperada para uma associação aleatória entre as densidades de lagostins e a presença de peixes.

A questão agora é de que tamanho deve ser *DIF* para que nós rejeitemos a hipótese nula e continuemos a acreditar em nossa hipótese alternativa? Para responder isto, precisamos calcular *DIF* quando a hipótese nula é "verdadeira" (os Popperianos ortodoxos diriam: quando a hipótese nula *não é falsa*). Para isto, vamos usar uma moeda e atribuir aleatoriamente a qualidade de "peixes presentes" (cara) e de "peixes ausentes" (coroa) aos valores medidos de densidades de lagostins, constantes da tabela 4. Criamos as variáveis PEIXE1, PEIXE2 e PEIXE3 da tabela 4 por este sorteio, com a restrição de que somente cinco riachos de cada tipo (PEIXE + ou PEIXE -) foram admitidos, para que, como na amostra original, houvesse sempre o mesmo número de riachos com e sem peixes.

As colunas PEIXE e LAGOSTINS mostram os dados usados para construir a figura 24a. As colunas PEIXE1, PEIXE2 e PEIXE3 foram o resultado do sorteio e são alocações ao acaso da presença ou ausência de peixes. Podemos parear cada uma destas colunas com a coluna das densidades "observadas" de lagostins (LAGOSTINS) para construir gráficos dos resultados esperados quando a hipótese nula é verdadeira (figuras 25a, b, c).

TABELA 4

RIACHO	PEIXE	LAGOSTINS	PEIXE1	PEIXE2	PEIXE3
1	+	1	+	+	+
2	-	5	-	+	-
3	+	3	+	+	+
4	-	7	-	-	-
5	+	4	+	-	-
6	-	8	+	-	+
7	+	5	-	-	-
8	-	9	+	-	+
9	+	6	-	+	+
10	-	9,5	-	+	-

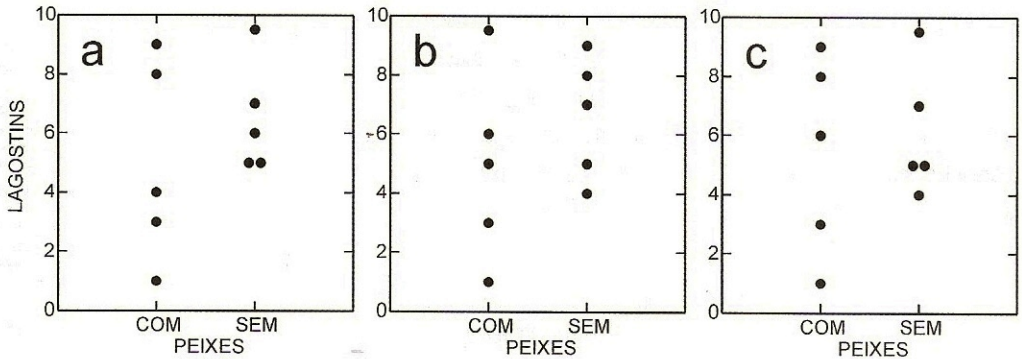
FIGURA 25

Agora, podemos calcular *DIF* para cada um destes gráficos. O primeiro tem uma *DIF* = -1,5, o segundo tem uma *DIF* = -1,7 e o terceiro uma *DIF* = -0,7. Se tivéssemos feito mais desses gráficos de resultados esperados quando a hipótese nula é verdadeira, qual seria a probabilidade de encontrarmos uma diferença tão grande ou maior que a calculada para os dados "observados" (*DIF* = -3,9)? Para responder a isto, precisaríamos de mais um grande número de gráficos e de lançar a moeda muitas e muitas vezes – uma maneira ineficiente de alocar a característica de presença ou ausência de peixes para os riachos. Felizmente, um computador pode eficientemente simular este processo que acabamos de descrever. Então, pedimos a ele para calcular 100 *DIF*s baseadas na alocação aleatória da presença ou ausência de peixes em riachos. A tabela 5 traz as primeiras 20 *DIF*s calculadas pelo computador.

TABELA 5

GRÁFICO	DIF	GRÁFICO	DIF
1	2,70	11	2,1
2	1,90	12	2,7
3	-0,09	13	-3,1
4	2,50	14	1,5
5	-2,70	15	-1,1
6	1,70	16	-0,1
7	0,50	17	1,3
8	0,10	18	-0,5
9	1,50	19	-0,7
10	3,10	20	-0,3

Todos os primeiros 20 resultados tiveram um valor absoluto menor do que o valor da *DIF* observada de -3,9, mas ainda restaram muitas combinações possíveis entre as densidades de lagostins observadas e o atributo de presença ou ausência de peixes nos riachos. Quando comparamos a *DIF* observada com as 100 *DIF*s calculadas sob a hipótese nula (figura 26), encontramos dois casos que tiveram uma diferença absoluta tão grande ou maior do que o valor observado.

FIGURA 26

Portanto, concluímos que há somente em torno de duas chances em 100 de se obter um valor absoluto de *DIF* tão grande ou maior que 3,9, quando não há associação entre a presença de peixes e a densidade de lagostins em riachos. Consideramos os valores absolutos dos resultados porque fizemos a pergunta geral se peixes afetam ou não a densidade de lagostins. Este é um teste de duas caudas. Se tivéssemos perguntado se a presença de peixes faz diminuir a densidade de lagostins, poderíamos fazer um teste mais sensível, de apenas uma cauda, contando somente quantos resultados simulados eram iguais ou menores a -3,9. Entretanto, o princípio geral é o mesmo e não queremos nos alongar em detalhes.

Vamos rever nossos passos até aqui, porque o processo é geral para todos os testes inferenciais. Processos similares são advogados por muitos autores (p. ex. Huberty 1987).

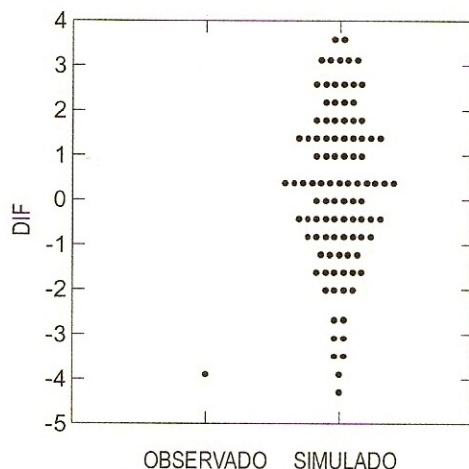
Passo 1. Visualize o resultado esperado quando a hipótese que você está testando for a correta.

Passo 2. Visualize o resultado esperado quando a hipótese que você esta testando não for a correta. Este resultado deve ser a antítese do anterior, ou seja as hipóteses devem ser opostas e excludentes. Ele é chamado de hipótese nula e este é o processo de criação da hipótese nula.

Passo3. Crie uma medida que reflita a diferença que você espera entre a situação em que a hipótese nula seja a correta e quando a sua hipótese (hipótese alternativa) seja a correta. Esta medida é chamada de uma "estatística".

Passo 4. Obtenha muitos valores de sua estatística, usando um processo no qual a hipótese nula seja verdadeira (*i.e.* sua hipótese seja falsa).

Passo 5. Compare o valor da estatística para os dados observados com os valores da estatística calculados quando a hipótese nula é verdadeira (valores nulos). Então use a proporção de valores nulos iguais ou maiores que o valor observado como uma indicação da probabilidade de se obter um valor tão grande ou maior do que o observado, quando a hipótese nula é verdadeira.

Passo 6. Decida se a probabilidade é suficientemente pequena para que você rejeite a hipótese nula e continue acreditando em sua hipótese (note que você nunca prova que sua hipótese está correta, apenas que a hipótese nula provavelmente está errada).

Se está parecendo complicado, retorne à figura 24. Deve ser suficiente para clarificar. Será uma boa idéia anotar este processo, já que o usaremos em todo o decorrer deste livro. Tínhamos uma questão complexa, que envolvia uma centena de gráficos. Usamos alguma matemática simples para calcular uma medida na qual estávamos interessados (neste caso, uma diferença) e usamos os valores resultantes para construir um único gráfico simples, que nos possibilita tomar uma decisão. Mesmo a lógica intrincada da filosofia Popperiana pode ser representada em um simples gráfico!

Deve estar claro, agora, porque os cientistas procuram por probabilidades pequenas. Eles estão se perguntando se a ciência conhecida, ou o senso comum, estão errados, e vão inventando explicações novas de como o mundo funciona. Basicamente, eles aceitam que podem estar certos em suas novas explicações, quando a probabilidade das explicações tradicionais estarem certas for muito pequena. Para mostrar que as explicações tradicionais estão erradas, eles precisam criar um processo (hipótese nula) que represente a maneira como a ciência conhecida ou o senso comum espera que o mundo funcione. Grande parte das polêmicas na literatura científica se deve a desavenças sobre qual é a hipótese nula, como medi-la e o quanto as evidências disponíveis são suficientes para rejeitá-la. Um bom delineamento amostral usualmente tem uma única interpretação lógica. Delineamentos amostrais inadequados levam a muitas interpretações alternativas e poucas conclusões podem ser tiradas com confiança.

Como os livros texto contam a estória?

A estória que contamos é simples, uma vez que você passe a usar a lógica Popperiana. Entretanto, os livros de estatística fazem parecer que algo muito complexo foi feito. Parte do problema é histórico, e parte é devido a inversão desnecessária de gráficos. Você não precisa se preocupar com a próxima seção deste capítulo, a menos que já tenha tido algum contato com a estatística e deseja entender como nossos gráficos se relacionam com os que são usualmente mostrados nos livros.

Para compreender as diferenças, vamos considerar a análise mais usada para se determinar quando duas amostradas são diferentes, o teste *t* de Student. O nome real de Student foi William S. Gosset e ele trabalhou em uma cervejaria. Sua hipótese nula era um pouco diferente da que usamos, quando fizemos o teste sorteando a moeda. A pergunta de Gosset foi "Qual é a probabilidade de que as duas amostras foram tiradas ao acaso da mesma população de medidas?" A lógica desta questão é óbvia. Duas amostras tiradas ao acaso de uma mesma população só podem ser diferentes por acaso. Gosset questionou-se "Com que freqüência esperaríamos amostras com médias tão ou mais diferentes do que a de nossas amostras se elas realmente vieram da mesma população de medidas?".

A hipótese nula de Gosset é similar a nossa, ao perguntar se algum outro processo, além do acaso, foi provavelmente responsável pelo resultado de uma diferença entre as médias tão grande ou maior que a dos valores observados. Entretanto, Gosset percebeu que a diferença entre as médias depende da escala na qual medimos as variáveis. Medições em milímetros

resultam em diferenças numéricas 10 vezes maiores que medidas em centímetros. Portanto, um conjunto diferente de *DIF*s esperadas sob a hipótese nula precisaria ser calculado para cada escala de medidas, e seria complicado a um tempo comparar coisas pequenas como camundongos e coisas grandes como antas.

Para resolver o problema da escala de medidas, ele padronizou as diferenças, dividindo-as pelo seu desvio padrão (reveja o capítulo 3 se você esqueceu o que é um desvio padrão), para deixar os resultados em uma mesma escala. Sua nova fórmula ficou como DIF/DP_{DIF}. Por agora, não se preocupe a respeito do desvio padrão da diferença (DP_{DIF}), basta que compreenda que foi um truque matemático para resolver o problema da escala. A fórmula de Gosset DIF/DP_{DIF} ficou mais tarde conhecida como a estatística *"t"* [veja Mosteller e Tukey (1968) para outras implicações históricas do teste de Gosset].

A vida não era fácil no início do século XX. Lembre-se que usamos um computador para calcular para nós as 100 *DIF*s, porque o processo de lançar moedas e calcular uma nova *DIF* para cada amostra de hipótese nula era muito demorado. Gosset não tinha acesso a um computador veloz, e precisou simular o processo escrevendo 3000 números em pedaços de cartolina, embaralhá-los e retirar 750 amostras de quatro unidades (Student 1908). Isto foi adequado para ilustrar a distribuição geral de sua estatística. Entretanto, ele considerou que seu gráfico não estava satisfatório, porque arredondara os números quando os escreveu nos pedaços de cartolina. A idéia de escrever outros 3000 números com muitas casas decimais não agradou muito Gosset e, rapidamente, compreendeu que coletar fisicamente amostras sob a hipótese nula era muito complicado, para ser usado em muitas situações. A partir daí ele embarcou numa viagem de "e se" que só um matemático poderia vislumbrar.

Se conhecermos alguma coisa a respeito da distribuição de medidas em nossa população nula hipotética, e se a distribuição destas medidas tiver uma forma que permita que a teoria matemática trabalhe com ela, poderíamos saber qual a distribuição que a estatística *t* deveria ter quando a hipótese nula estivesse correta. Ele usou os dados coletados para estimar algumas características da população nula hipotética. Características de populações são chamadas de parâmetros. Por isso, este tipo de teste foi chamado de paramétrico.

Como Gosset não teve facilidades de computação para gerar muitas distribuições de valores de *t* quando a hipótese nula era correta, usou a teoria matemática para estimar a proporção de *t*s hipotéticos tão grandes ou maiores que o *t* observado. O brilhantismo de Gosset é revelado pelo fato de que, quando suas premissas a respeito da distribuição da população hipotética são satisfeitas, seu método leva a um resultado extremamente próximo aos resultados gerados pelos métodos de computação intensiva, similares ao que usamos.

Como os estatísticos contam observações independentes?

Quando a hipótese nula está correta, uma população diferente de *t*s existe para cada combinação de número de observações em cada amostra. Nós, comparamos cinco riachos com peixes e cinco riachos sem peixes. Entretanto, se comparássemos sete riachos com peixes e cinco sem peixes, teríamos encontrado diferentes populações de *DIF*s e de *t*s. Isto não muda muita coisa em nosso método de computação intensiva, talvez um milionésimo de segundo a mais para fazer as computações adicionais. Entretanto, sem a ajuda de computadores velozes, o processo é proibitivo em termos do tempo gasto, e os estatísticos do início do século XX precisaram inventar o conceito de valores críticos para as estatísticas. Agora eles podiam dizer: "Se você tem 5 observações na primeira amostra e 7 na segunda, todos os valores de *t* maiores do que 1,81 terão menos do que 10% de chance de serem observados, quando a hipótese nula estiver correta. Valores de *t* maiores do que 2,95 terão menos do que 1% de chance".

Estes valores críticos foram usados para se construir as tabelas de valores críticos usualmente encontradas no final dos livros de estatística. A maioria dos programas de computadores, especializados em estatística, podem calcular a probabilidade exata da hipótese nula ser verdadeira, e as tabelas estatísticas quase não são mais usadas. Contudo, este exemplo introduz o conceito de graus de liberdade usados nos testes estatísticos. Para encontrar a probabilidade associada com qualquer valor em particular de uma estatística, precisamos conhecer quantas observações independentes foram usadas para calculá-la. Usualmente se diz que os graus de liberdade dos testes estatísticos são o número de observações independentes menos o número de parâmetros estimados. Em nosso exemplo, tínhamos 10 riachos e estimamos a média e o desvio padrão, para construir a distribuição de resultados esperados quando a hipótese nula era verdadeira. Portanto, o grau de liberdade para a estatística *t* em nosso exemplo foi 10 - 2 = 8.

Mas, para interpretar corretamente as tabelas estatísticas, precisamos na verdade de dois números, um que se refere ao número de parâmetros estimados e outro ao número de observações independentes. Será mais consistente com outros testes se considerarmos os graus de liberdade de nosso exemplo como sendo 1 e 8. No caso de nossa comparação de duas amostras, não precisamos nos preocupar em apresentar o primeiro número, porque o teste *t* sempre estima apenas 2 parâmetros. Não é preciso preocupar-se com os detalhes. Por enquanto, basta notar que os graus de liberdade são associados com o número de observações independentes. Se suas observações não forem independentes (*i.e.* você cometeu pseudo-repetição), você entrará na tabela no ponto errado e a probabilidade que estimar será uma pseudo-probabilidade, que não pode ser relacionada com a hipótese nula em questão. Infelizmente, muitas (senão a maioria) das probabilidades apresentadas na literatura são pseudoprobabilidades, que só servem para indicar que o autor pertence a cultura científica, e não transmitem qualquer informação objetiva sobre o mundo real.

Não se preocupe se achou esta última seção de difícil leitura. O importante aqui é entender que estatísticos como Gosset seguiram o mesmo processo do nosso teste de cara-ou-coroa.

Passo 1. Ele visualizou o resultado esperado quando sua hipótese de como o mundo deveria funcionar era correta.

Passo 2. Visualizou o resultado esperado quando sua hipótese não era correta.

Passo 3. Ele criou uma medida que refletia a diferença esperada entre a situação quando a hipótese nula era a correta e quando sua hipótese era a correta. Esta medida é agora conhecida como a estatística *t*.

Passo 4. Ele usou a teoria matemática para obter muitos valores da estatística usando um processo no qual a hipótese nula era verdadeira.

Passo 5. Ele usou a matemática para comparar o valor da estatística estimada para os dados observados com a estatística esperada quando a hipótese nula era verdadeira (valores nulos). Então, usou a proporção esperada de valores nulos iguais ou maiores que o valor observado como um indicador da probabilidade de se obter um valor daquela magnitude ou maior quando a hipótese nula fosse verdadeira.

Passo 6. Ele decidiu se essa probabilidade era tão pequena que ele deveria rejeitar a hipótese nula e continua a acreditar em sua hipótese.

É mais fácil compreender o mundo se não você virá-lo de cabeça para baixo

Se o leitor leu alguns livros de estatística e se lembra de coisas a respeito da distribuição normal e outras distribuições, será tentado a pular diretamente para a próxima seção. Mas, por favor, não faça isto, porque **o mundo ficará desnecessariamente complexo se nos afastarmos de nossos gráficos simples de dispersão.** Uma razão pela qual os estudantes ficam confusos com as distribuições estatísticas e com valores de dados é que a distribuição de valores de uma medida ou uma estatística é sempre apresentada na horizontal nos livros de estatística. Para entender isto, vamos inverter alguns gráficos. Considere a distribuição dos dados de pontos na figura 27a.

Relacionamos os valores de densidades de lagostins com categorias de riachos. Porque consideramos as densidades de lagostins como sendo dependentes das categorias de riachos, e não o contrário, colocamos as densidades de lagostins no eixo vertical (eixo Y), seguindo a convenção de que a variável dependente é representada no eixo Y. Entretanto, se tivéssemos apenas uma categoria de riachos, como nas hipóteses nulas de testes paramétricos, a única coisa que poderia variar seria a freqüência de cada densidade dos lagostins. Por causa disto, as distribuições são convencionalmente apresentadas com a variável dependente no eixo horizontal (eixo X). Neste caso, os dados da figura 27a passariam a ser apresentados como na figura 27b.

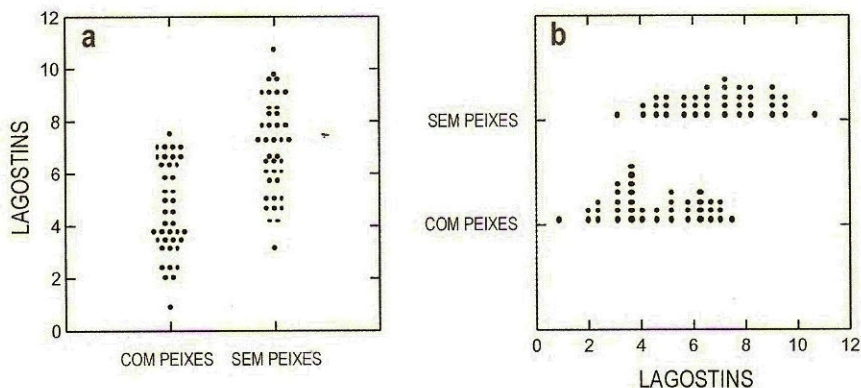
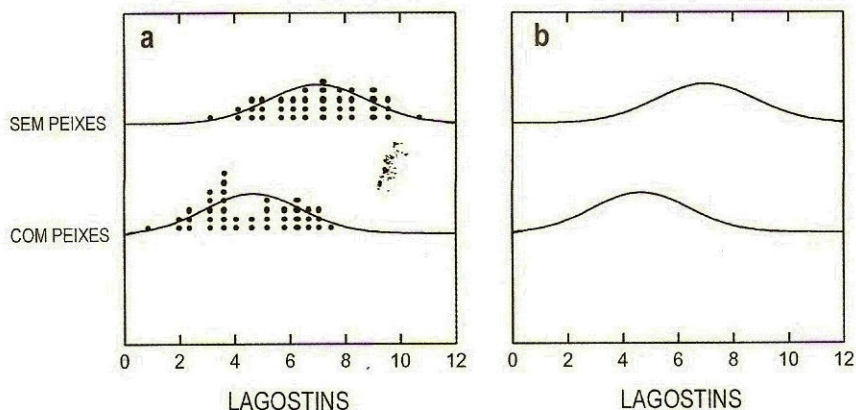
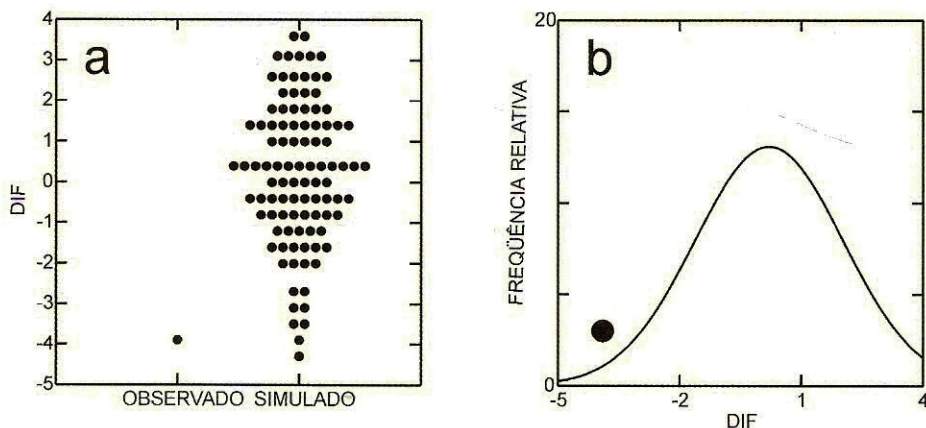
FIGURA 27

Entretanto, os matemáticos não vêem pontos individuais e, assumindo uma distribuição em particular chamada de "distribuição normal", usam os dados para estimar as características desta distribuição (figura 28a) e descartam os dados, ficando apenas com a curva teórica (figura 28b).

FIGURA 28

Podemos ver que a distribuição normal nem é uma aproximação muito boa dos dados dos riachos com peixes, mas *geralmente isto não afeta a validade dos testes estatísticos comumente usados. A distribuição de dados esperados quando a* hipótese nula é verdadeira é, na verdade, uma única curva híbrida que é um tipo de média matemática das duas curvas da figura 28b. Usamos esta curva para estimar a diferença nas médias para as amostras de um determinado tamanho, quando a hipótese nula está "correta".

O tipo de distribuição esperado é similar ao da nossa *DIF* (figura 26). Para comparar nossos valores observados de *DIFs* com a distribuição esperada de *DIFs* quando a hipótese nula é "verdadeira" (figura 29a - a mesma da figura 26), poderíamos inverter os eixos, apagar os pontos e comparar apenas a posição da bolinha preta com a distribuição teórica de *DIF* ao longo do eixo X (figura 29b).

Isto é o que a maioria dos livros de estatística nos ensina a fazer. Entretanto, isto não ajuda muito na compreensão do processo. Os matemáticos não amostram fisicamente suas populações nulas, mas se imaginarmos que eles fazem esta amostragem e se tentarmos imaginar as distribuições das estatísticas no eixo vertical, ficará mais fácil relacionar o processo com os nossos próprios dados. Quase todas as análises que vamos considerar destinam-se a examinar a variação dos valores expressos nos eixos verticais de nossos gráficos (variáveis dependentes). Portanto, nós apenas complicaríamos mais as coisas se insistíssemos em inverter a orientação dos gráficos.

FIGURA 29

Tudo isto está ficando um pouco complicado, portanto, vamos retornar aos exemplos do mundo real. No próximo capítulo, vamos considerar como podemos testar eficientemente diferenças entre várias categorias.

[Lidamos com probabilidades e coisas improváveis acontecem.]

Como evitar acumular o risco em comparações simples

Com que tipo de risco estamos preocupados?

O fato de constatarmos que um determinado resultado não é freqüente quando a hipótese nula está correta não quer dizer que nunca o observaremos quando ela estiver correta. Lidamos com probabilidades e coisas improváveis acontecem. A probabilidade de sua sogra morrer atropelada, hoje mesmo, é muito pequena. Entretanto, milhares de sogras morrerão atropeladas esta noite.

O fato das probabilidades com que os estatísticos usualmente trabalham serem muito diferentes daquelas com que as pessoas se preocupam no dia-a-dia, sempre causam dificuldades aos estudantes. Isto acontece, principalmente porque os cientistas tentam evitar decidir que um fenômeno existe quando ele não existe. Isto faz sentido, porque a ciência avança sobre o conhecimento anteriormente acumulado. Se este conhecimento é falso, tudo o mais que for baseado nele estará errado. Se decidíssemos que peixes afetam a densidade de lagostins, quando na verdade não afetam, vamos deixar de ver a causa limitante real e tomaríamos medidas erradas para aumentar a densidade de lagostins (caso isto fosse desejável).

Tradicionalmente, os cientistas da área biológica têm considerado que um fenômeno existe se há menos do que uma chance em 20 (0,05) de que ele não exista. Esta probabilidade que funciona como um critério para se determinar se o fenômeno existe ou não (neste caso 0,05), é chamada pelos estatísticos de "nível de significância" (ou em seus hieróglifos, "δ"). Vemos que os cientistas não abraçam novas explicações facilmente. [Rejeitar erroneamente a hipótese nula e decidir que um fenômeno existe, quando ele não existe, é chamado de erro do tipo I.] A estatística convencional é construída para se resguardar deste tipo de erro.

Entretanto, erros do tipo I nem sempre são os mais custosos. Isto é mais facilmente compreendido com um exemplo pessoal. Imagine que você se apaixone por um(a) estatístico(a), e desejem ter filhos. Imagine que ele ou ela tenha justamente retornado de uma conferência onde tenha se envolvido em relacionamentos promíscuos. Entretanto, seu(sua) parceiro(a) argumenta: "Não se preocupe, eu fiz todos os cálculos e a chance de eu *não* ter contraído alguma doença sexualmente transmissível é de 6%, e a probabilidade de 0,06 não é suficiente para rejeitar a hipótese nula de que eu não tenha sido contagiado(a)". Foi uma conclusão válida do ponto de vista estatístico e científico. Contudo, provavelmente você não se importará muito em cometer o erro do tipo I, porque a probabilidade de aceitar a hipótese nula quando ela é falsa é muito mais custosa, neste caso.

[Aceitar a hipótese nula, quando ela é falsa, é chamado de erro do tipo II e a probabilidade se cometer um erro do tipo II é conhecida como "δ" pelos estatísticos.] Mesmo que haja apenas uns 20% de chance de seu parceiro ser portador de uma doença contagiosa e potencialmente fatal, ainda não será suficiente para que você continue seu relacionamento sem uso de preventivos e, possivelmente, até probabilidades muito menores não serão suficientes para que você leve adiante a idéia de filhos, sem exigir previamente dele(a) uma bateria de exames laboratoriais.

Para uma amostra de determinado tamanho, a probabilidade de se cometer o erro do tipo II é inversamente proporcional à probabilidade de se cometer o erro do tipo I. Por isso, muitos cientistas usam níveis de significância muito maiores do que 0,05 quando o erro do tipo II envolve um custo alto. Isto ocorre comumente em estudos que tratam de saúde humana, extinção de espécies, e quando a rejeição prematura da hipótese levará ao abandono de uma linha de pesquisa potencialmente muito importante. Poucos cientistas aceitariam o nível de significância arbitrário de 0,05 para planejar sua vida pessoal. Entretanto muitos deles estão prontos para aceitá-lo em suas atividades profissionais, refletindo a falta de seriedade que devotam a ciência. A capacidade de se detectar uma diferença quando ela realmente existe (isto é, de não se cometer um erro do tipo II) é chamado de "poder" do teste.

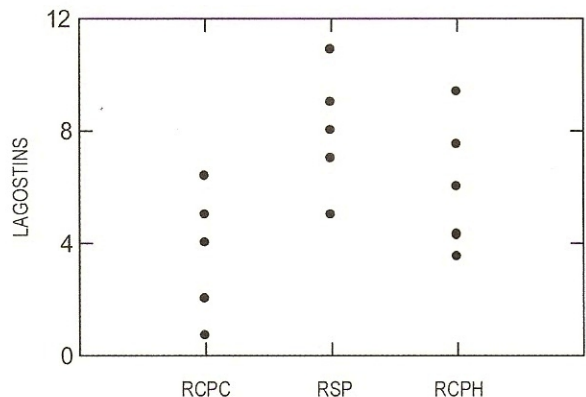
Neste livro, não teremos tempo para enfatizar a importância de erros do tipo II e para a maioria das discussões que se seguirão assumiremos que estamos lidando com situações nas quais o erro do tipo I é o mais custoso. Isto facilitará o aprendizado dos conceitos gerais. Entretanto, esperamos que os leitores não se esqueçam de avaliar a importância de erros tipo II em sua pesquisa. Referindo-se especificamente a uma análise, Koele (1982) declarou que "Uma avaliação adequada dos resultados experimentais sob estes modelos é simplesmente impossível sem o conhecimento a respeito do poder dos testes que usam a estatística *F*". Diríamos que este comentário se aplica a todos os modelos e todos os testes estatísticos e não apenas àqueles que usam a estatística *F*. Referências úteis sobre este tópico incluem Koele (1982), Huberty (1987) e Green (1989). Informando-se alguns dados preliminares, alguns programas de computador podem calcular o tamanho amostral necessário para se detectar efeitos de magnitude e probabilidades especificados (p.ex. SIMSTAT, Péladeau, 1996). Entretanto, sempre haverá alguns "truques" matemáticos envolvidos nestas operações e geralmente será melhor o leitor estar certo de ter alguns gráficos hipotéticos convincentes.

Usando a variabilidade para reconhecer uma diferença

No capítulo anterior, consideramos um exemplo no qual comparamos uma amostra com outra e avaliamos a probabilidade de declarar erroneamente que alguma coisa além do acaso causou a diferença entre elas, construindo uma hipótese nula. Não faz diferença se realizamos uma permutação física (nosso teste das *DIFs*) ou usamos uma matemática complicada (teste *t* de Student) para gerar os resultados esperados quando a hipótese nula estivesse correta. Os princípios são os mesmos. Entretanto, nossa hipótese nula se relacionou somente a comparação de dois grupos. Vamos considerar agora o que acontece quando comparamos mais do que dois grupos. Talvez nosso biólogo esteja interessado nas densidades de lagostins em riachos sem peixes (RSP), riachos com apenas peixes herbívoros (RCPH) e riachos com peixes carnívoros (RCPC). Seus resultados são mostrados na figura 30.

Poderíamos usar nosso teste de cara-coroa ou o teste *t* de Student para comparar RSP com RCPH, RSP com RCPC, e RCPH com RCPC. Entretanto, cada vez que executarmos o teste corremos o risco de cometermos um erro. Por enquanto, vamos seguir a convenção e assumir que só devemos rejeitar a hipótese nula, de que não há diferenças entre as categorias de riachos, se houver apenas uma chance em 20, ou menos, disso acontecer ao acaso. Se nada além da sorte estiver causando as diferenças nas médias, somente rejeitaremos a hipótese nula quando 5% (0,05) ou menos das *DIFs* simuladas (ou *ts* calculados) forem menores do que a *DIF* (ou *t*) dos dados observados. Portanto, a chance de cometermos o erro do tipo I quando comparamos RSP com RCPH é cerca de uma em 20. Entretanto, se formos comparar RSP com RCPC, novamente, teremos uma chance em 20 de encontrarmos uma diferença, quando ela não existe. A probabilidade geral de, por puro azar, encontrarmos uma diferença que não existe, agora é de duas em 20. Com três testes será aproximadamente três em 20 e continua aumentando, embora não linearmente, com o número de testes.

Esta taxa cumulativa de erro nos é familiar em nosso dia-a-dia, embora, quando jovens, somos inclinados a crer que nunca acontecerá conosco. A maioria dos pais ficam receosos quando seus filhos pegam o carro emprestado e dirigem em alta velocidade, ou praticam esportes radicais como vôo livre, ou simplesmente não têm o bom senso de praticar sexo seguro. Toda vez que uma criança vai contra as orientações de seus pais, se torna mais e mais convencida de que é imune aos perigos. Entretanto, os pais sabem que se um comportamento de risco é repetido um número suficiente de vezes, uma pequena chance de perigo se torna quase uma certeza. Os cientistas também tendem a encolher os ombros e não dar atenção para o problema de testes repetidos, porque eles estão convencidos de que suas hipóteses estão corretas. Geralmente são

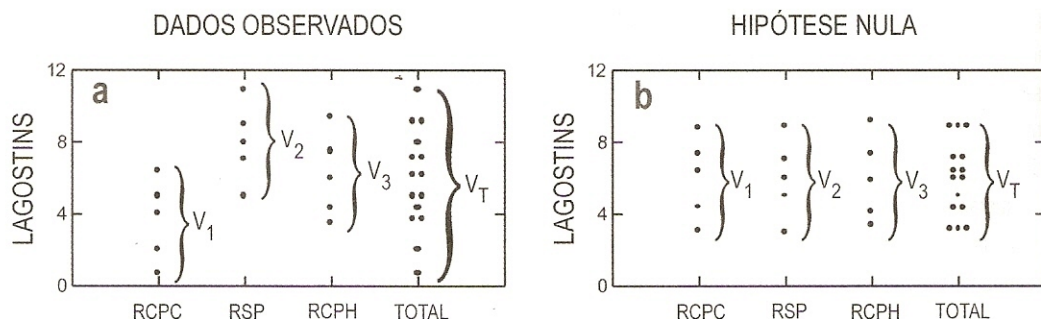
FIGURA 30

necessárias algumas experiências desagradáveis até que eles se convençam de que as leis da probabilidade se aplicam a todas as pessoas, inclusive cientistas ! Tukey (1991) apresenta uma discussão iluminada sobre a filosofia de múltiplas comparações, que se relaciona com questões do mundo real.

Bonferroni sugeriu, há muito tempo, que quando fazemos mais de um teste para a mesma hipótese, podemos corrigir a probabilidade de rejeitar a hipótese nula geral, de ausência de diferenças entre as amostras, multiplicando a probabilidade de cada teste pelo número de testes realizados. Se qualquer das probabilidades multiplicadas for pequena o suficiente para rejeitar a hipótese nula, então a hipótese nula geral é rejeitada. Em nosso caso, decidimos só rejeitar a hipótese nula se houver no máximo uma chance em 20 de ela estar correta. Portanto, rejeitaremos a hipótese nula se qualquer das probabilidades calculadas nos testes, após serem multiplicada por três, for menor do que que o nível de significância escolhido de 0,05.

A correção de Bonferroni é simples, fácil de fazer, e tão boa quanto qualquer outro método quando poucos testes estiverem sendo feitos. Entretanto, no caso de muitos testes, o procedimento de Bonferroni tende a aceitar a hipótese nula mais do que deveria (erro do tipo II). Modificações neste procedimento podem incrementar seu poder (Rice 1989, Benjamini e Hochberg 1995), mas algumas vezes também mudam a questão que está sendo respondida (Benjamini e Hochberg 1995).

Ronald Fisher desenvolveu um método superior para comparar muitas categorias. O problema de usarmos *DIF* ou *t* é que um gráfico simples como o da figura 30 tem muitos *DIF*s ou *t*s. Fisher raciocinou que seria mais eficiente se o gráfico pudesse dar origem a apenas uma estatística que refletisse a diferença geral entre as categorias. Fisher foi um matemático e podia raciocinar em termos de diferenças nos desvios quadrados (reveja o capítulo 3 se achar que precisa saber mais alguma coisa sobre desvios quadrados). Podemos examinar a lógica de Fisher usando a amplitude em lugar das diferenças nos desvios quadrados, uma vez que a amplitude é facilmente reconhecível em gráficos (figura 31).

Incluímos uma outra categoria, combinando os dados de todos os tipos de riachos e a chamamos de variabilidade total (V_T). Se a variabilidade dentro de cada categoria é a mesma e a média for diferente, então a variabilidade total será

FIGURA 31

muito maior que a de cada categoria. Embora tenhamos usado a amplitude (V) para representar a variabilidade das categorias, a mesma lógica se aplica com qualquer outra medida de variabilidade, incluindo a variância, que foi a medida usada por Fisher. Em termos matemáticos, quando a hipótese nula não é verdadeira, como na figura 31a, essa relação pode ser representada pela equação a seguir.

$$V_1 = V_2 = V_3 < V_T$$

Esta mesma equação pode ser escrita como $\bar{V}_i < V_T$, lembrando que a barra sobre o V_i significa que se trata de uma média e que o índice subscrito "$_i$" está substituindo, de forma genérica, os índices que representam as categorias 1 a 3. Se equações lhe dão arrepios, recorra aos gráficos para apreciar a mesma informação.

A figura 31b mostra um resultado esperado quando a hipótese nula está correta. Não há diferença entre as médias e, neste caso, podemos ver que a variabilidade dentro da categoria "variabilidade total" é igual a variabilidade dentro de cada categoria. Podemos expressar isto matematicamente pela equação a seguir:

$$V_1 = V_2 = V_3 = V_T$$

que também poderia ser expresso como $\bar{V}_i = V_T$. Esperamos que possam ver isto no gráfico, embora o gráfico para este exemplo pareça um pouco forçado (figura 31b). Nele, a variabilidade em cada categoria foi quase igual a das demais e as médias foram exatamente iguais. É claro que a chance disto ocorrer com amostras reais é desprezível. Entretanto, não esperamos diferenças muito grandes quando a hipótese nula é "verdadeira" e, neste sentido, o gráfico preenche nossas expectativas.

A comparação dos dois gráficos mostra que, quando a hipótese nula é "verdadeira", esperamos que a variação média dentro de cada categoria seja aproximadamente igual a variação total. Uma estatística simples que reflete isto é a variação total dividida pela variação média dentro das categorias (*i.e.* $V_T / \bar{V}_i$). Vamos chamar nossa estatística conceitual de "*RV*" (razão da variação). Quando a hipótese nula for "verdadeira", $V_T = \bar{V}_i$ e $RV = 1$. Quando a hipótese nula é falsa, $V_T > \bar{V}_i$ e $RV >> 1$. Usamos dois símbolos de "maior que" ($<$) em seguida para indicar que *RV* é muito maior do que um. A idéia é essa. No entanto, assim como Student não usou a estatística *DIF*, Fisher não usou *RV*. Ele usou a razão de duas variâncias. Fisher chamou sua estatística de razão de variâncias, mas em sua homenagem ela foi chamada de estatística *F*. A estatística *F* de Fisher é construída de uma forma um pouco diferente de nossa *RV*, como veremos na seção de partilha da variabilidade, mas assim como em nossa *RV*, o valor esperado para *F* quando a hipótese nula é "verdadeira" é sempre 1. Como sempre, se as equações deixaram-no confuso, retorne a figura 31 e certifique-se que pode enxergar a diferença entre as situações quando a hipótese nula não é verdadeira (figura 31a) e quando ela é "verdadeira" (figura 31b). As equações e as estatísticas apenas refletem a diferença mostrada no gráfico.

Para gerar amostras de *F* quando a hipótese nula era "verdadeira", Fisher teve de lançar mão de uma matemática pesada. Da mesma forma que Student, Fisher teve que assumir que a distribuição de valores na população nula hipotética

seguia a distribuição normal, embora os resultados de seu teste não tenham sido muito sensíveis a desvios moderados desta premissa. Se o leitor estiver especialmente preocupado com a forma da distribuição da população da qual seus dados foram tomados, sempre poderá usar testes de permutação como o nosso usando *DIF*. Manly (1997) apresenta muitos exemplos de testes de permutação e outros assemelhados, que não requerem todas as premissas dos testes paramétricos.

Uma premissa importante

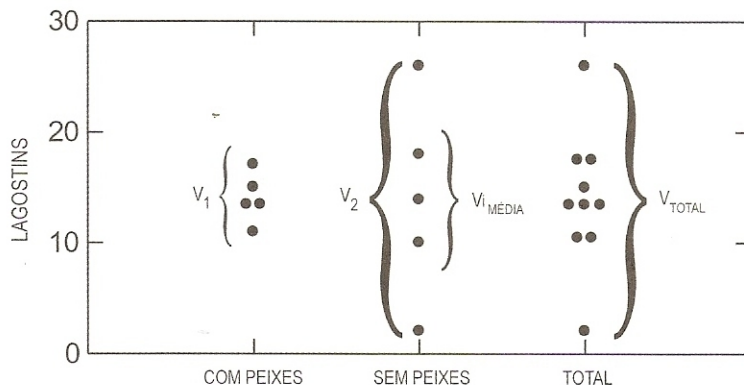
As lógicas dos testes de Fisher e de Student são muito parecidas. De fato, quando se compara apenas duas amostras, $F = t^2$. Já dissemos que eles são relativamente robustos em relação à premissa de "normalidade" da população da qual os dados são tirados. Entretanto, há uma outra premissa que não pode ser deixada de lado. Grandes valores de *F* podem indicar ou que as médias são muito diferentes entre as "categorias" (níveis do fator), ou que a variabilidade dentro das categorias difere entre as categorias. Se desejamos testar as diferenças entre médias, precisamos assumir que a variabilidade dentro das categorias é aproximadamente igual.

[Se desejamos testar as diferenças entre médias, precisamos assumir que a variabilidade dentro das categorias é aproximadamente igual.]

Tanto a estatística *F* quanto a *t* são calculadas dividindo-se a variabilidade entre as médias das categorias pela variabilidade dentro das categorias. A lógica é que se as amostras foram tomadas de uma única população com uma determinada variação média (hipótese nula), esperamos encontrar diferenças entre as médias na mesma proporção prevista nas tabelas estatísticas ou pelos computadores. Entretanto, a figura 32 mostra um exemplo onde a variabilidade diferiu entre as categorias (presença ou ausência de peixes). O computador não enxerga a variabilidade de cada categoria (V_1 e V_2) e cria a distribuição nula baseada na variabilidade média ($\bar{V}_i$). Se as amostras representadas na figura 32 representam a variabilidade do universo de cada categoria de riachos, então podemos perceber que seria muito mais fácil encontrar valores extremos para médias de amostras tiradas da distribuição representada pela categoria de riachos sem peixes (com grande variabilidade) do que para a distribuição representada pela média de V_i. Portanto, a distribuição nula não está correta e as probabilidades tenderão a ser pequenas, indicando diferenças nas médias onde elas não existem. Se aplicarmos o teste de Fisher (chamado de análise de variância ou ANOVA, a sigla em Inglês para "*Analysis of Variance*"), para testar por diferenças nas médias entre os níveis do fator, provavelmente descobriríamos uma diferença significativa, mas esta diferença seria devida a subestimativa da variabilidade na hipótese nula. Isto nos levaria a cometer um erro do tipo I, *i.e.* estaríamos rejeitando a hipótese nula de ausência de diferença entre as médias quando a hipótese nula era verdadeira.

Embora a maioria dos livros de estatística se preocupe com a homogeneidade das variâncias (ou seja, se as variâncias são aproximadamente iguais), o teste pode ser mais sensível no que diz respeito a simetria. A análise de variância e muitos testes análogos de aleatorização são muito sensíveis a pontos muito extremos, chamados pelos estatísticos de "*outliers*". Assim, é muito importante inspecionar o gráfico antes de aplicar qualquer avaliação estatística.

Erros do tipo II são comumente encontrados em testes para determinar se as premissas de uma análise estão sendo cumpridas. O teste de Levene, por exemplo, é usado para determinar se as variâncias dentro dos níveis do fator não são

FIGURA 32

suficientemente homogêneas, o que resultaria em valores improváveis de *F*. O uso de baixos valores críticos para a probabilidade (P) é geralmente tido como "seguro" ou "conservador". Entretanto, para testes de premissas, o erro do tipo I não é o mais custoso. Se você erroneamente deixar de identificar que a premissa está incorreta, atribuirá um efeito significativo onde ele não existe. Os autores de livros de estatística freqüentemente recomendam um valor crítico de P = 0,05 para estes testes (p. ex. Dytham 1999). Se a probabilidade calculada foi de 0,07, isto equivale a escrever que "há 7% de chance de que minha conclusão não seja completamente sem sentido". Isto parece "seguro" para você?

[Por alguma razão os biólogos têm fixação em comparar médias. Entretanto, freqüentemente, diferenças na variabilidade são mais importantes (p. ex. Callaghan e Holloway 1999). Daqui para diante, iremos sempre discutir a variabilidade, embora algumas vezes usaremos a variabilidade para fazer inferências sobre médias ou outras características (parâmetros) de populações.]

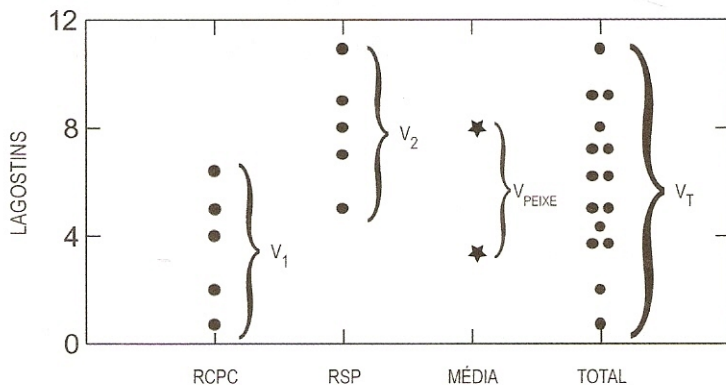
A análise de variância nos diz que há uma diferença entre as médias, porém não nos diz quais médias são diferentes. Para determinar isto, é preciso olhar o gráfico ou usar um teste mais fraco, como o teste de Tukey, para localizar a diferença (p. ex. Day e Quinn 1989). Não vamos nos preocupar com estes testes agora, já que se o leitor compreender os conceitos por detrás de testes múltiplos e ANOVA, será capaz de entender as comparações entre tratamentos após uma ANOVA.

Partição das variâncias

Antes de terminarmos com o modelo de análise de variância de Fisher, vamos considerar alguma terminologia, que será útil quando tratarmos de exemplos mais complexos. Vamos considerar os dados mostrados na figura 33, no qual a variabilidade dentro das categorias é similar entre as categorias de riachos com peixes (RCP) e riachos sem peixes (RSP). Temos a variabilidade dentro das categorias (V_1 e V_2), as médias das categorias e a variabilidade total (V_T).

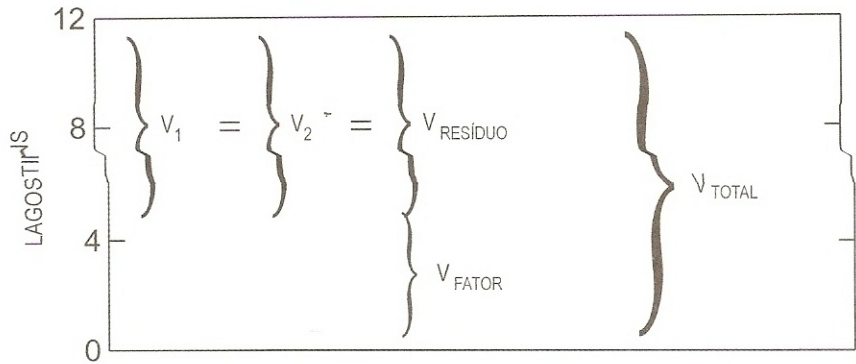
Conceitualmente, podemos dizer que a variabilidade entre as categorias (V_{PEIXE}) decorreu da presença de peixes e que a presença ou ausência de peixes é o "tratamento" ou "fator" que potencialmente está afetando a densidade de lagostins. A diferença entre a variabilidade total e a V_{PEIXE} é devida a variabilidade residual, que não pode ser atribuída a uma causa em particular. A variabilidade residual é algumas vezes chamada de "erro", mas isto implica que esta variabilidade não faz parte do mundo real, o que não é verdade.

Se fizermos um processo análogo, mas usando a variância ao invés da amplitude para representar a variabilidade, estaremos fazendo o que os estatísticos chamam de "partição das variâncias". Os cálculos ficam um pouco mais complicados, mas o princípio é o mesmo. Este é o processo usado pela maioria das análises complexas que estaremos examinando no resto deste livro. Portanto, se o leitor ainda tem dificuldades com o processo de partição da variabilidade em gráficos, por favor, reveja esta seção. Vamos continuar nos referindo a partição da variabilidade como mostrada na figura 33 e os leitores só serão capazes de acompanhar nossa linha de raciocínio se visualizarem este processo em termos de dados exibidos em gráficos.

FIGURA 33

Experimentem para ver se são capazes de visualizar a seguinte equação simples nas figuras 33 e 34.

$$V_{FATOR} + V_{RESÍDUO} = V_{TOTAL}$$

A figura 34 é basicamente igual a figura 33. Mas afim de deixá-la um pouco mais universal, em relação aos conceitos que desejamos transmitir, introduzimos pequenas modificações. Para facilitar a representação da variação

FIGURA 34

residual ($V_{RESÍDUO}$) que não aparece explicitada na figura 33, nós alinhamos V_1 e V_2 e abstraímos os pontos. Na figura 33 a variação devido ao fator "presença de peixes" aparece como V_{PEIXE}, enquanto na figura 34 a variação devido ao fator aparece como V_{FATOR}.

Fisher trabalhou com uma quantidade que ficou conhecida como "soma dos quadrados" (*SQ*) e tomou decisões baseadas nas "médias dos quadrados" (*MQ*), mas não é preciso saber calcular estas coisas para entender os princípios por detrás da análise. Ao invés de comparar a variação residual com a variação total, a estatística *F* de Fisher é a razão entre a média dos quadrados do fator (tratamento) e a média dos quadrados dos resíduos. Na ANOVA, a média dos quadrados do fator não representa apenas a variabilidade entre as médias. Ela é uma estimativa da variabilidade entre as médias mais a variabilidade dentro dos níveis do fator ou tratamento (*i.e.* variação residual).

As médias dos quadrados são calculadas como a soma dos quadrados dividida pelos graus de liberdade. Portanto, elas são análogas às variâncias que calculamos no capítulo 3. Entretanto, elas podem ser variâncias compostas. A média dos quadrados do fator é uma estimativa da variância devido ao fator mais a variância residual, e a média dos quadrados do resíduo é uma segunda estimativa da variância do resíduo. Conceitualmente, e esquecendo umas poucas constantes, temos a seguinte equação:

$$F = (\delta^2_{Fator} + \delta^2_{Resíduo})/\delta^2_{Resíduo}$$

onde δ^2 significa a variância. Quando a variância devido ao fator for zero (a hipótese nula está "correta"), $F = 1$. Como foram baseadas em variâncias compostas, as razões *F* podem se tornar complicadas, como veremos no capítulo 8. Entretanto, o conceito de que a variabilidade pode ser repartida em uma fração devido ao fator e outra devido à variação residual é simples.

Os estudantes sempre perguntam como os programas de ANOVA podem calcular $F < 1$, se a equação do parágrafo anterior for correta. Isto acontece porque uma mesma quantidade conceitual, a variância do resíduo, é estimada de formas diferentes no numerador e no denominador da equação. Devido às incertezas da amostragem e porque o valor esperado pelos estatísticos é diferente do valor que nós, simples mortais, esperamos encontrar, ocorre do resultado das estimativas serem algumas vezes diferentes. Na verdade, a maioria dos *F*s são menores do que um, quando a hipótese nula está "correta".

O valor esperado para um estatístico é o valor que se poderia esperar se repetíssemos o exercício um grande número de vezes e usássemos a média. Para uma distribuição enviesada para a direita (quer dizer, que têm a cauda mais espichada para o lado direito), como a dos valores de *F*, um valor ocasional muito elevado pode resultar em um *F* médio próximo ou igual a um, embora a maioria dos valores seja menor do que um. Para os estatísticos, valores esperados sem vieses, relacionam-se com o valor médio esperado em uma infinidade de tentativas. Já o ecólogo usualmente está tentando tomar uma decisão baseado em um único experimento.

As tabelas de *F* levam em conta o desvio para a direita e fornecem as probabilidades corretas. Entretanto, se o leitor estiver construindo seus próprios testes, compreenda que alguns valores "esperados" podem não estar tão próximos daqueles que você espera encontrar freqüentemente. Caughley and Sinclair (1994:210) apresentam um exemplo para a estimativa de Petersen de tamanho populacional que mostra como os valores esperados daquela estimativa são muito maiores que a maioria dos valores calculados.

Muitos cursos de estatística começam com exemplos de tabelas de contingência como a maneira mais simples de se analisar categorias. Entretanto, estas análises só são simples dentro de salas de aula e são virtualmente sem utilidade em estudos ecológicos (veja o exemplo do capítulo 4, Hurlbert 1984, ou Magnusson 2000b). Na maioria das vezes em que for necessário lidar com categorias, modelos de análise de variância serão mais simples e eficientes do que análises baseadas em tabelas de contingência.

CAPÍTULO 7

[As diferentes cores não são categorias, mas expressões de uma variável contínua]

Análises para um mundo com todas as tonalidades

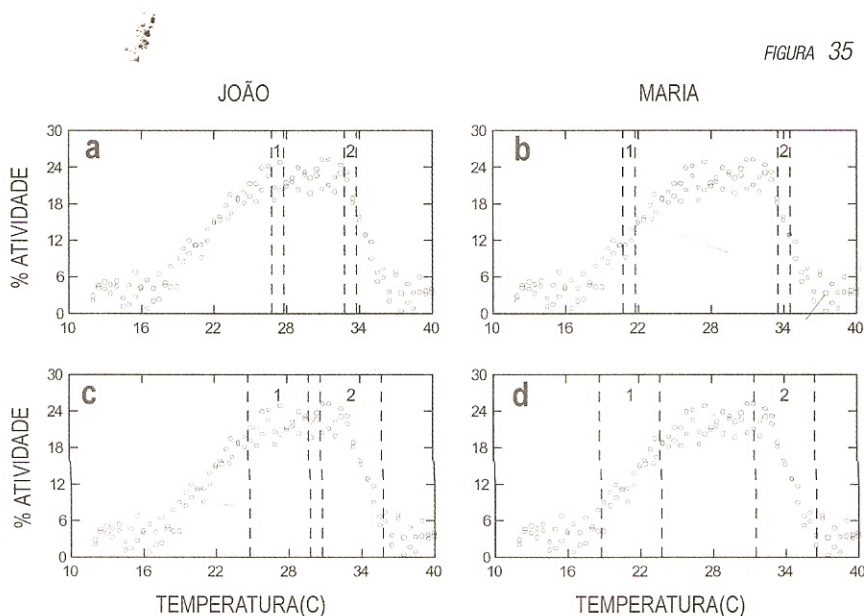
Passamos bastante tempo fazendo considerações de como testar as diferenças entre categorias. Em nossa vida cotidiana somos cautelosos em relação a pessoas que insistem em colocar tudo em categorias. Chamamos estas pessoas de sexistas, racistas, xenófobas ou outros termos que, quase sempre, carregam um tom de nossa censura ou desconfiança em relação ao seu modo de pensar. Usualmente, estas pessoas não são socialmente muito competentes (embora Hitler seja freqüentemente considerado como a figura política mais importante do século XX). Estudos científicos usualmente começam com categorizações, mas tendem a estagnar-se até que alguém comece a estudar os processos ao invés de apenas os padrões.

Para entender algo sobre as cores é preciso entender que o olho e o cérebro humano interpretam uma estonteante diversidade de cores a partir de apenas três tipos diferentes de receptores de cores em suas retinas. Os engenheiros de equipamentos de computação usam esta informação para projetar impressoras coloridas que reproduzem milhões de cores usando combinações de pigmentos vermelho, verde, azul e preto. De qualquer maneira, em última análise, a única coisa que realmente muda de uma cor para outra é o comprimento da onda eletromagnética, uma variável contínua.

Em nosso cotidiano, reconhecemos diferenças sexuais "fixas", mas esta categorização é grosseira para muitos propósitos, inclusive na medicina. A maioria das drogas não interage diretamente com o sexo do organismo, mas com níveis de hormônio, conteúdo de gordura do corpo, taxa metabólica, densidade dos ossos, e uma multidão de fatores que só de longe podem ser relacionados com a aparência da genitália. Mesmo no contexto social, a categoria sexo não funciona muito bem. Muitas pessoas homossexuais, com todo direito, recusam-se a ter sua sexualidade definida com base apenas em alguns detalhes anatômicos. Se as categorias são uma base fraca para nossas relações sociais, por quê elas têm uma posição tão exaltada na ciência?

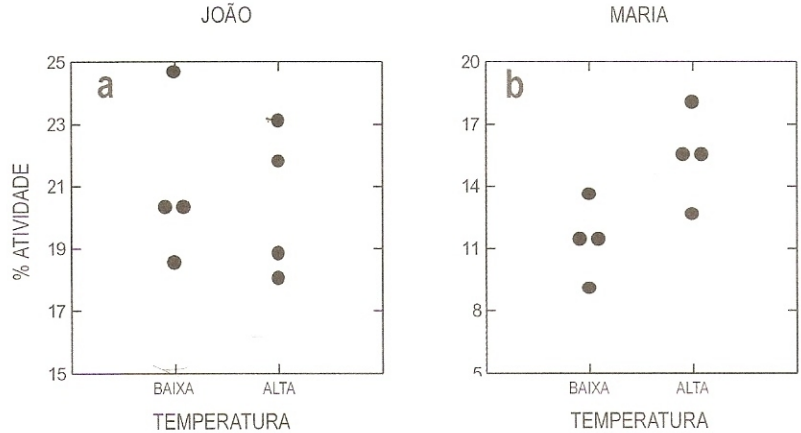
Encaixotando o mundo

Vamos considerar o que acontece quando categorizamos um fenômeno contínuo e depois investigar formas mais diretas de se atacar o problema. A figura 35 mostra dados sobre a atividade de um inseto em relação a temperatura do ar. Suponha que dois biólogos, João e Maria, decidiram fazer experimentos para investigar a atividade deste inseto em função da temperatura, e delinearam seus experimentos de forma a amostrar a temperatura em dois níveis: baixa (= 1) e alta (= 2). Entretanto, cada um decidiu usar níveis diferentes para representar "alto" e "baixo" e estes níveis estão mostrados na figura 35a e b. Notem que eles não tiveram informação sobre a atividade dos insetos fora dos níveis de temperatura que escolheram. Depois de coletarem seus dados, cada um produziu um gráfico para mostrar os resultados (figura 36).

FIGURA 35

Já que João e Maria leram o capítulo anterior deste livro, decidiram testar se as diferenças entre as médias das amostras em temperatura "alta" e "baixa" deveriam ser atribuídas ao acaso. Eles poderiam ter usado nosso teste de *DIF*, mas decidiram usar o teste *t* de Student, com uma correção para as diferenças entre as variâncias dentro de cada categoria. O leitor não precisa se preocupar com este refinamento, introduzido para garantir que a premissa de homogeneidade das variâncias do teste *t* não fosse violada, porque, no fim das contas, este teste vai dar resultados similares ao de simulações de *DIF*.

O teste de João estimou uma probabilidade de 0,78 de que a hipótese nula fosse correta. Conseqüentemente, João rejeitou sua hipótese de que a temperatura afeta a atividade do inseto. O teste de Maria estimou uma probabilidade de 0,035

de que a hipótese nula estivesse correta, portanto rejeitou a hipótese nula como improvável, aceitando que a temperatura afeta a atividade do inseto. A maioria das pessoas tenderão a chegar às mesmas conclusões, apenas examinando os gráficos da figura 36, dispensando o uso de qualquer estatística extravagante ou lógica Popperiana.

FIGURA 36

A obtenção destes resultados contraditórios é compreensível em termos de um modelo de análise de variância. Diríamos que toda a variabilidade devido a amostragem está na variabilidade residual. Este modelo conceitual poderia ser escrito como:

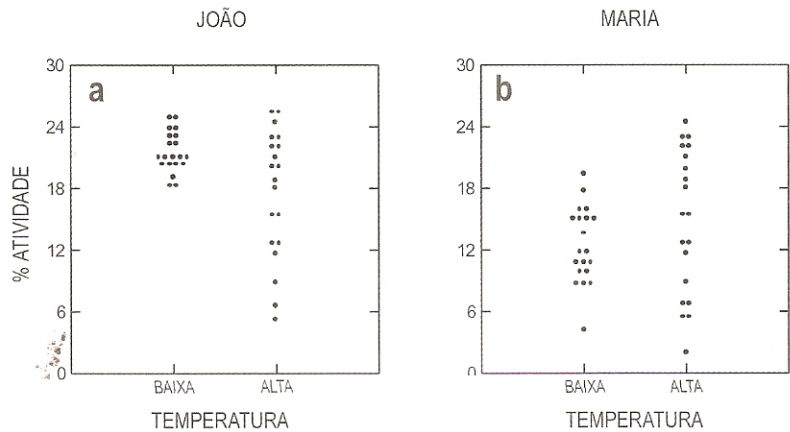
$$V_{RESÍDUO} + V_{FATOR} = V_{TOTAL}$$

Entretanto, no caso da amostragem de João e de Maria, há uma outra fonte de variação que afeta a variação total: a variação devido ao fato de terem amostrado apenas alguns dos níveis possíveis do fator e o modelo conceitual deve incorporar um termo para esta variabilidade. Podemos chamar este termo de $V_{NÍVEIS}$. Então, o modelo conceitual passa a ser:

$$V_{RESÍDUO} + V_{FATOR} + V_{NÍVEIS} = V_{TOTAL}$$

Quando amostramos todos os níveis de um fator, dizemos que este é um fator "fixo". Se amostramos apenas uma pequena proporção dos possíveis níveis do fator, ele é chamado de "randômico". Entretanto, nenhum destes termos é muito apropriado. Quase sempre, os níveis de um fator randômico são amostrados de uma forma não aleatória, e nem todos os possíveis níveis têm igual probabilidade de serem amostrados. Este delineamento inadequado pode afetar muito o resultado. Vamos discutir isto depois, porque agora nos concentraremos em outros efeitos resultantes de se categorizar variáveis contínuas, que tornam a interpretação dos resultados ainda mais difícil.

Voltando a figura 35, vamos supor que João e Maria poderiam ter escolhido níveis de temperatura "alta" e "baixa" que apresentassem as mesmas médias que as apresentadas nas figuras 35a e 35b, mas com intervalos mais largos para cada nível, como mostrado nas figuras 35c e 35d. Se fizermos os gráficos de seus resultados, exatamente como fizemos os gráficos

FIGURA 37

com níveis estreitos mostrados na figura 36, nossas conclusões serão opostas (figura 37). Usando o mesmo teste *t* de Student, João agora rejeita a hipótese nula como sendo improvável (P = 0,013) e Maria aceita a hipótese nula (P = 0,22).

A tabela 6 compara as conclusões de João e Maria quando usam intervalos estreitos ou largos para definir suas categorias. Nosso modelo conceitual agora indica que a variação total não apenas depende dos níveis amostrados, mas também da amplitude dos intervalos escolhidos para representar a categoria.

$$V_{RESÍDUO} + V_{FATOR} + V_{NÍVEIS} + V_{AMPLITUDE} = V_{TOTAL}$$

Existem algumas técnicas estatísticas para lidar com a variação devida aos níveis do fator ($V_{NÍVEIS}$) e vamos tecer considerações sobre elas no próximo capítulo. Entretanto, não há um método objetivo de lidar com a variação devida à escolha da amplitude destes níveis ($V_{AMPLITUDE}$).

TABELA 6

INTERVALOS	JOÃO	MARIA
Amplos	Significante	Não significante
Estreitos	Não significante	Significante

Neste ponto, esperamos que os leitores estejam se questionando porque os pesquisadores desprendem energia para categorizar variáveis contínuas. Parte da resposta é que os seres humanos se sentem confortáveis com categorias aparentemente simples. Como dissemos anteriormente, este é o motivo pelo qual tendemos a ser racistas, sexistas e xenófobos. Uma outra razão é que experimentos de "inferência forte" usualmente só são possíveis quando se usa categorias. Quando existe um número infinito de intervalos, não é possível assegurar que todos os níveis de cada categoria sejam medidos em todos os níveis de outro fator que também

esteja sendo estudado. Além disso, pode haver uma grande informação na ordem em que as "categorias" ocorrem na natureza, com a qual a análise de variância de categorias padrão simplesmente não consegue lidar (Gaines e Rice 1990).

Não se preocupem se não compreendem toda esta terminologia aqui; iremos retornar a ela posteriormente. Basicamente, quando um pesquisador categoriza uma variável contínua em nome do delineamento experimental, está trocando uma inferência fraca a respeito de uma categoria forte (há apenas uma categoria) por uma inferência forte a respeito de categorias subjetivas, e portanto, fracas. Isto está relacionado com nossas discussões sobre escalas nos capítulos anteriores (Magnusson 2002a) Hierarquias baseadas na escala (quantidades mensuráveis) podem ter uma base objetiva. Hierarquias baseadas em "níveis" sempre serão subjetivas e podem impedir o avanço da ciência (Allen 1998).

Descrevendo um mundo retilíneo

Para entender como computadores e testes estatísticos lidam com variáveis contínuas, precisamos começar com o tipo de relação contínua mais simples: uma linha reta. A figura 38 mostra um mapa de árvores (representadas pelo símbolo "o") em 6 reservas de tamanhos diferentes.

FIGURA 38

Vamos construir um modelo simples deste sistema. Podemos esperar uma relação direta entre a área da reserva e o número de árvores em cada reserva, se não houver outros fatores atuando. O modelo está representado pela linha na figura 39. Os pontos representam o número de árvores contados em cada reserva.

Nosso modelo teórico parece ser bastante bom, julgando-se pelo fato de que em geral os pontos estão perto da linha. O modelo teórico pode ser representado por uma equação e, usando o símbolo * para representar a multiplicação, teremos a seguinte forma:

NÚMERO DE ÁRVORES = A + B * ÁREA DA RESERVA

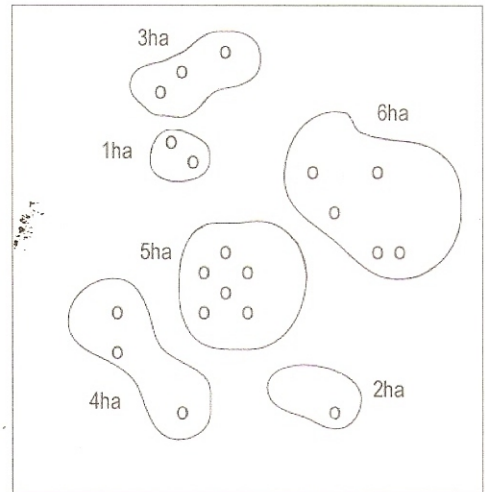
Os livros especializados usariam letras gregas para "A" e "B", porque elas são teóricas e não foram estimadas a partir dos dados. Em nosso exemplo, A = 0 e B = 1. Substituindo na equação, veremos que a equação diz que quando

não há reserva (área = 0), não há árvores. A equação diz também que para cada hectare extra, teremos, em média, uma árvore a mais. Esta equação geral pode ser usada para descrever uma linha reta. O "A" é a elevação, ou o valor da variável dependente quando a variável independente é igual a zero. O "B" é a inclinação, ou a quantidade que a variável dependente aumenta para cada unidade de aumento da variável independente. Já que a maioria das análises freqüentes em ecologia é baseada em variações desta equação, vale a pena empregar algum tempo e se certificar de que podem visualizar esta equação no gráfico.

FIGURA 39

Entretanto, na maior parte das vezes, não estamos tratando com equações teóricas e tentaremos determinar a posição da linha a partir dos dados coletados. A figura 40 mostra uma relação entre a atividade de um inseto e a temperatura, na faixa entre 26 e 36º C. Não conhecemos a verdadeira relação entre a atividade do inseto e a temperatura, mas esboçamos no gráfico a linha que achamos que provavelmente representa esta relação, baseados apenas na distribuição dos pontos.

[Quando nós, outras pessoas, ou programas de computadores colocam linhas retas em gráficos, fazem isto minimizando a distância média da linha aos pontos.] Isto parece uma ação simples e direta, mas na verdade há muitas maneiras diferentes pelas quais as distâncias podem ser minimizadas.

FIGURA 40

A figura 41 mostra uma parte de nosso gráfico, e apresenta três maneiras de como as distâncias podem ser minimizadas.

A figura 41a mostra uma forma lógica de minimizar a distância de dois pontos até a reta. A operação de minimizar as distâncias da reta até os pontos, como mostrado na figura 41a, é o que faz o método chamado de regressão do maior eixo. Já a regressão do "maior eixo reduzido" encontra a posição da reta que resulta na menor área dos triângulos formados pelas linhas horizontais e verticais do

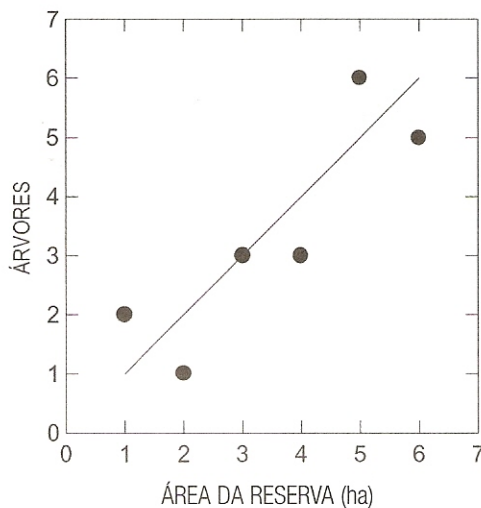
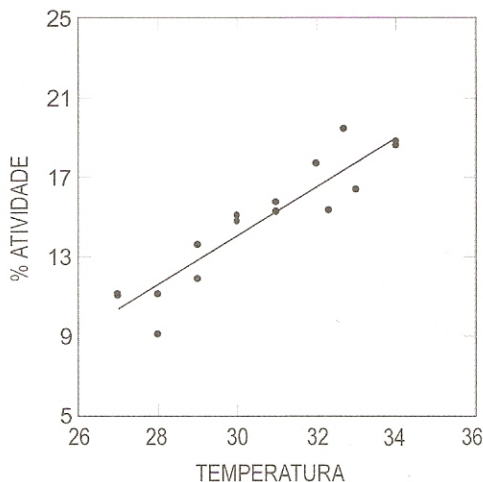
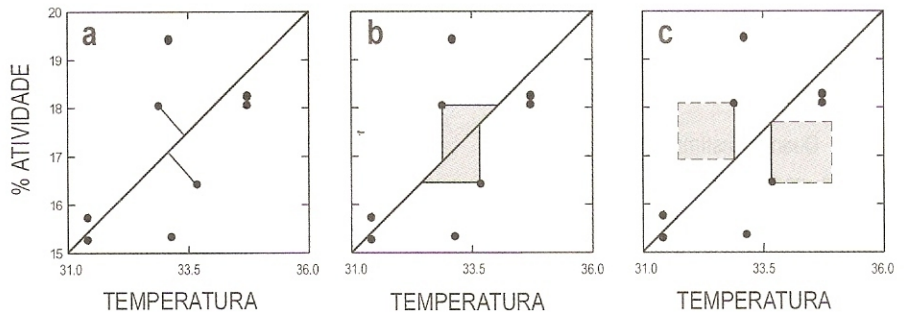



FIGURA 41

ponto até a reta (figura 41b). Estes dois métodos são lógicos e provavelmente os mais intuitivos para a maioria das pessoas. Minimizar apenas a distância vertical (linha contínua na figura 41c) não parece muito lógico. Minimizar a soma dos quadrados das distâncias verticais dos pontos até a reta (a área contida nos quadrados na figura 41c), parece ainda mais ilógico. Entretanto, este é o processo por trás do método estatístico mais empregado. Cada um destes métodos de minimizar as distâncias tem vantagens e desvantagens. Não vamos nos ater neste ponto e se os leitores desejarem uma discussão a respeito, podem consultar Ricker (1973).

A minimização dos quadrados das distâncias verticais até a linha permitiu o desenvolvimento de uma grande variedade de análises complexas e geralmente as mais úteis em termos de previsão. Por razões óbvias (veja figura 41c), ela é chamada de "regressão dos mínimos quadrados". Ela é matemática e logicamente equivalente a análise de variância de Fisher, para variáveis categóricas independentes. As principais variações das análises estatísticas mais comuns se resumem a minimizar diferentes distâncias. Vamos fazer algumas considerações sobre isto no capítulo 11. Estas distâncias são resíduos, da mesma forma que a variação dentro das categorias na análise de variâncias.

Quanto o modelo se ajusta?

A variação em torno da linha é a variação resídual não explicada pelo modelo (a linha) e é ilustrada na figura 42. Quando a variabilidade é medida pela soma de quadrados (veja capítulo 3), a proporção da variabilidade que pode ser atribuída a variável independente (neste caso a temperatura) é chamada de coeficiente de determinação e é simbolizada como r^2 ou, se há mais de uma variável independente, R^2 (capítulo 8).

Embora provavelmente a maioria dos leitores terá dificuldade em visualizar a soma de quadrados (imagine a soma das áreas dos quadradinhos pontilhados da figura 41c), certifique-se de que possa visualizar o tipo de variação representado na figura 42 como V_{FATOR}/V_{TOTAL}. Para análises que envolvem categorias, temos que assumir que a variação dentro das categorias é constante. Para regressões, precisamos assumir que a variação ao redor da linha é constante ao longo de toda a linha. Embora

o r^2 seja uma das estatísticas mais usadas, não é uma boa medida da magnitude do efeito em análises categóricas (Rosenthal and Rubin 1982).

FIGURA 42

Quando estimamos a posição da linha a partir dos dados, usamos letras minúsculas para representar a elevação e a inclinação na equação descritiva Y = a + b*X, onde Y é qualquer variável dependente (em nosso caso "% atividade"), e X é qualquer variável independente (no caso, "temperatura"). As letras "a" e "b" representam estimativas de parâmetros (estatísticas), porque descrevem a população de pontos nos quais a regressão foi baseada. A hipótese nula pode ser representada na situação mostrada na figura 43.

Da mesma forma que nas análises categóricas, quando a variação residual é tão grande quanto a variação total, assumimos que não há efeito daquela variável independente sobre nossa variável dependente, ou $V_{FATOR} = 0$. Na verdade, o modelo de análise de variância é geral e a maioria dos programas de computadores executam a análise de variância de dados categóricos e regressão exatamente da mesma forma. Portanto, tanto para regressão quanto para análise de variância, nosso modelo conceitual começa exatamente pela mesma simples equação:

$$V_{FATOR} + V_{RESÍDUO} = V_{TOTAL}$$

FIGURA 43

Na regressão, não precisamos nos preocupar com seleção de categorias, porque temos apenas uma categoria. Entretanto, precisamos nos preocupar com a amplitude do intervalo da categoria. Se a relação for realmente linear, a probabilidade de se detectar uma relação aumenta com o aumento da amplitude do intervalo. A figura 44 mostra como a relação forte mostrada na figura 42 desaparece quando diminuímos

Testes univariados são geralmente redundantes

Problemas do mundo real: mais de um fator

Embora os exemplos de testes estatísticos apresentados nos capítulos anteriores tenham sido úteis para ilustrar alguns princípios, é importante compreender que estes testes univariados (com apenas uma variável independente) são geralmente redundantes. Eles podem ser substituídos, com vantagens, por simples gráficos de dispersão. Pesquisadores inexperientes podem se deixar enganar por ilusões estatísticas (Anscombe 1973). Se a análise estatística sugere uma conclusão diferente do padrão que aparece no gráfico, não acredite em nenhum dos dois. Muitos tratamentos estatísticos são triviais e não servem a outro propósito que expressar identidade cultural (Yoccoz, N. G. 1991; Cherry 1999, Johnson 1999), ou seja, comunicar que a pessoa que está apresentando aquele resultado é um cientista.

Considere o fluxograma da figura 45. Se a variável independente (árvores) é categórica e tem mais do que dois níveis, a seta pode representar apenas o sentido da influência. Ela não pode representar que tipo de efeito a variável sofrerá. Por exemplo, se "árvores" está representando "espécies de árvores", podemos imaginar que o aumento de algumas espécies será bom para os macacos, mas o aumento de outras pode ser ruim ou ainda os macacos podem ser indiferentes a quantidade de indivíduos de muitas espécies de árvores.

FIGURA 45

ÁRVORES → MACACOS

Teremos, também, o mesmo problema se a relação é contínua, mas não linear, ou pelo menos, monotônica. Um aumento em uma variável contínua, como densidade de árvores, pode causar aumento do número de macacos quando

Testes univariados são geralmente redundantes

Problemas do mundo real: mais de um fator

Embora os exemplos de testes estatísticos apresentados nos capítulos anteriores tenham sido úteis para ilustrar alguns princípios, é importante compreender que estes testes univariados (com apenas uma variável independente) são geralmente redundantes. Eles podem ser substituídos, com vantagens, por simples gráficos de dispersão. Pesquisadores inexperientes podem se deixar enganar por ilusões estatísticas (Anscombe 1973). Se a análise estatística sugere uma conclusão diferente do padrão que aparece no gráfico, não acredite em nenhum dos dois. Muitos tratamentos estatísticos são triviais e não servem a outro propósito que expressar identidade cultural (Yoccoz, N. G. 1991; Cherry 1999, Johnson 1999), ou seja, comunicar que a pessoa que está apresentando aquele resultado é um cientista.

Considere o fluxograma da figura 45. Se a variável independente (árvores) é categórica e tem mais do que dois níveis, a seta pode representar apenas o sentido da influência. Ela não pode representar que tipo de efeito a variável sofrerá. Por exemplo, se "árvores" está representando "espécies de árvores", podemos imaginar que o aumento de algumas espécies será bom para os macacos, mas o aumento de outras pode ser ruim ou ainda os macacos podem ser indiferentes a quantidade de indivíduos de muitas espécies de árvores.

FIGURA 45

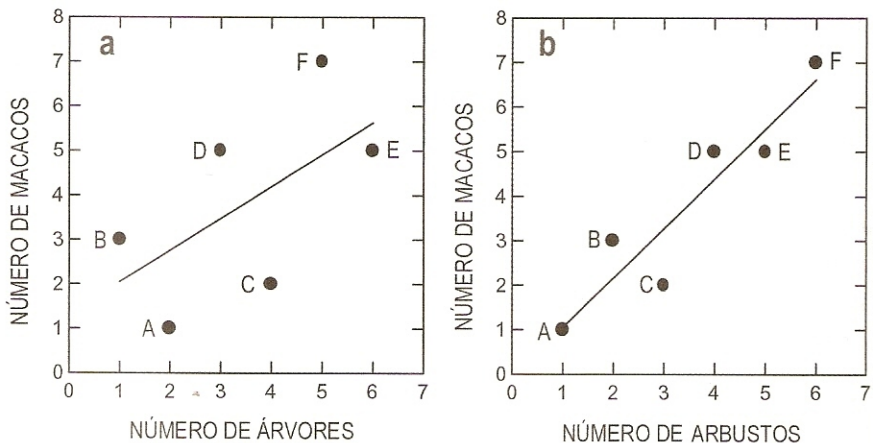
ÁRVORES ⟶ MACACOS

Teremos, também, o mesmo problema se a relação é contínua, mas não linear, ou pelo menos, monotônica. Um aumento em uma variável contínua, como densidade de árvores, pode causar aumento do número de macacos quando

as árvores são esparsas, mas um aumento semelhante pode determinar o decréscimo do número de macacos, quando a cobertura arbórea for muito densa. A figura 12 (capítulo 2) que apresenta a variação da atividade de um inseto em função da temperatura, ilustra este tipo de relação, muito freqüente na biologia. Vamos começar com um modelo mais simples, que não ocorre com freqüência na natureza, mas que é muito comum em modelos estatísticos. Vamos assumir que há uma relação linear entre árvores e macacos.

Fatores simultâneos

O gráfico que ilustra a relação entre macacos e árvores (figura 46 a) parece ser linear e mostra um aumento na densidade de macacos com a densidade de árvores. Esta relação parece lógica, porque macacos se refugiam de predadores terrestres em árvores. Se coletarmos dados sobre a densidade de uma espécie de arbusto que os macacos usam como alimento, também encontraremos uma relação linear positiva (figura 46 b). Estes dois gráficos sugerem o fluxograma apresentado na figura 47.

FIGURA 46

Baseados nestes gráficos, os estudantes geralmente sugerem que plantar árvores e arbustos deve ser um bom método de aumentar o número de macacos. No entanto, considerando os dados disponíveis, esta medida é imprópria para o objetivo proposto. Se analisarmos os dados pela técnica de regressão de mínimos quadrados, podemos produzir gráficos de dispersão

que ilustram os efeitos parciais de arbustos independente de árvores e de árvóres independente de arbustos. Para saber como isto é feito, primeiro temos de dar uma olhada nos quatro gráficos intermediários (figura 48).

FIGURA 47

As letras maiúsculas junto dos pontos indicam a reserva de onde vieram aqueles dados. As linhas verticais indicam o desvio de cada ponto em relação à linha, ou seja, a variabilidade não explicada pelo modelo linear. Na figura 48 a, a variabilidade de macacos explicada por arbustos é representada pela linha e os desvios indicam a variabilidade de macacos não associada a arbustos. Da mesma forma, podemos obter a variabilidade de árvores não associadas com arbustos na figura 48 b. As distâncias dos pontos até a linha, que chamamos desvios ou resíduos, refletem as mudanças no número de árvores ou macacos devido a outros fatores que não os arbustos.

Se esboçarmos o gráfico destes resíduos contra o número de arbustos, veremos que eles não indicam qualquer relação com os arbustos. Alguns autores diriam que eles descrevem a variabilidade esperada se arbusto fosse mantido constante. Se construirmos o gráfico dos resíduos da figura 48 a contra os resíduos da figura 48 b, pareando cada ponto pela reserva que o originou, teremos a relação entre macacos e árvores esperada se removêssemos o efeito de arbustos, ou, de forma equivalente, a relação esperada se o número de arbusto fosse constante entre as reservas (figura 49 a). A figura 49 a é a "regressão parcial" de macacos com árvores. Ela é chamada regressão parcial porque mostra apenas a parte da variabilidade nos dados que não está associada com arbustos.

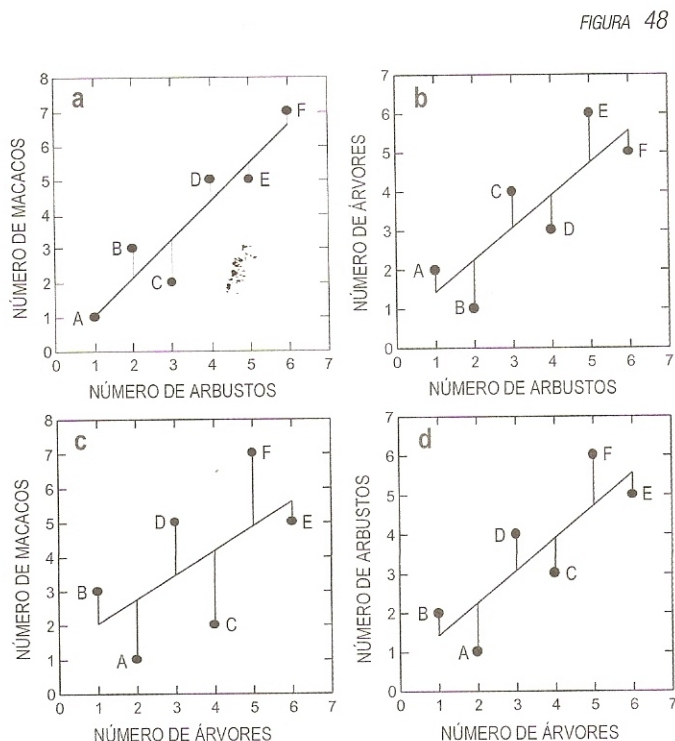
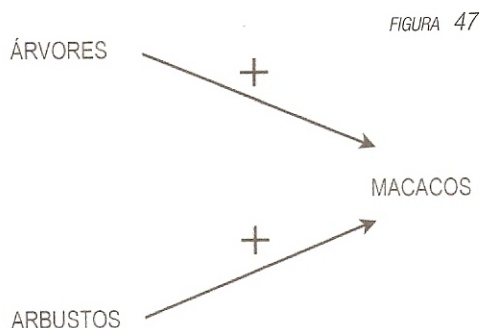
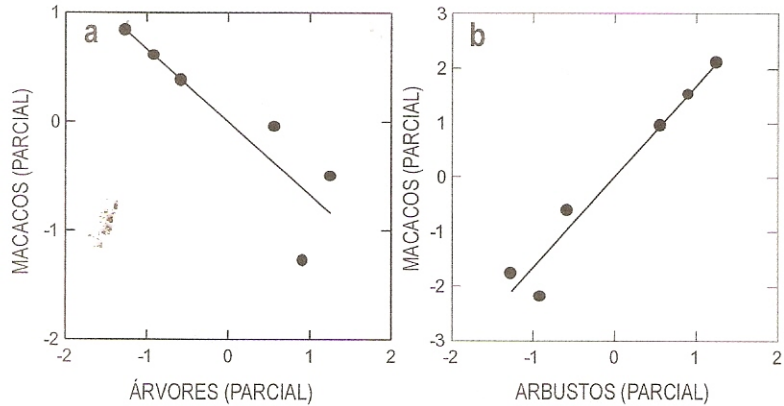



FIGURA 48

Podemos generalizar este processo e usar os resíduos das figuras 48c e 48d para produzir o gráfico da regressão parcial de macacos e arbustos, independentes de árvores (figura 49b). De fato, o método de regressão múltipla usa o mesmo processo para isolar os efeitos prováveis de cada fator independente dos efeitos de todos os outros fatores que aparecerem no modelo.

FIGURA 49

Adicionando variabilidades repartidas

Este processo de partilha da variação entre muitos fatores parece complicado no início, mas não é uma complicação imposta apenas por razões culturais. Ele nos ajuda a ver mais claramente. A relação simples entre macacos e árvores (figura 46a) é positiva, levando a maioria dos pesquisadores a acreditar que as árvores provavelmente eram boas para estes macacos. Entretanto, a regressão parcial (figura 49a) indica um efeito negativo de árvores sobre os macacos, quando estatisticamente mantivemos constante o efeito de arbustos. Na ausência de outras informações, certamente não deveríamos embarcar em um programa de plantio destas árvores, se o nosso objetivo for conservar ou aumentar o número de macacos. De fato, se a hipótese sugerida na figura 49a for corroborada, talvez devêssemos começar a cortar árvores como medida para aumentar o número de macacos.

O processo apresentado neste exemplo é geral e ilustra a forma pela qual a análise de variâncias ou outras técnicas de modelagem podem ser usadas para revelar padrões, que não estavam óbvios nos gráficos simples de dispersão originais. O processo implica na produção de outros gráficos de dispersão. Matematicamente, reduzimos o modelo descritivo da forma $Y = a + b_1 * X_1 + b_2 * X_2 + e$ para dois modelos da forma $Y_{PARCIAL} = a + b * X_{PARCIAL} + e$.

Notem que estes gráficos parciais são o verdadeiro resultado da análise, e não o sumário estatístico ou valores de probabilidade que os pacotes estatísticos jogam nas telas dos computadores. As estatísticas podem ser usadas como uma

informação acessória aos gráficos parciais, mas nunca poderão substituí-los. O leitor pode ter achado nosso exemplo trivialmente simples. Entretanto, muitos problemas complexos do mundo real têm sido enfrentados com sucesso por modelos de partilha de variâncias.

A equação para a regressão múltipla e regressões parciais nos contam a mesma estória. Empregando a técnica de regressão de mínimos quadrados, encontraremos a seguinte equação geral para o nosso exemplo:

$$\text{macacos} = 0{,}33 - 0{,}667 * \text{árvores} + 1{,}667 * \text{arbustos}$$

As regressões parciais são:

$$\text{macacos(parcial)} = 0 - 0{,}667 * \text{árvores(parcial)}$$

$$\text{macacos(parcial)} = 0 + 1{,}667 * \text{arbustos(parcial)}$$

A regressão geral tem as mesmas inclinações (valores de b) das regressões parciais. De fato, ela foi construída pela soma das regressões parciais. A elevação não é a mesma, porque um único valor não poderia substituir as elevações individuais de cada regressão parcial, que em nosso exemplo foram iguais (ambas as regressões passaram pela origem dos eixos *i.e.* a = 0), mas que poderiam ser diferentes. De qualquer modo, a similaridade da regressão geral com as regressões parciais ilustra sua natureza comum. A regressão múltipla não passa da soma de efeitos lineares estimados pelas regressões parciais. Este é um modelo simples, mas que provou ser útil em muitas situações.

Checando premissas com gráficos de parciais

Pesquisadores freqüentemente calculam a probabilidade de que o efeito de uma variável independente (valor de b) seja igual a zero (ou algum outro valor), estabelecendo assim sua hipótese nula. Não entraremos em detalhes aqui, mas os princípios de testar uma hipótese nula de b = 0 são similares a testar a hipótese nula de *DIF* = 0. Estas probabilidades podem ser calculadas por testes de randomização ou testes paramétricos, como o teste *t*. Entretanto, é importante fazer o teste paramétrico sobre a regressão múltipla geral, e não diretamente sobre as parciais. O motivo disto se deve ao fato de que o computador não sabe quantos outros fatores, e portanto, parâmetros, foram usados para calcular cada regressão parcial. Já que o número de graus de liberdade é calculado pelo número de observações independentes menos o número de parâmetros estimados, o computador usará o número errado de graus de liberdade em suas contas, caso não seja adequadamente informado sobre o número de parâmetros que foram estimados.

[De qualquer maneira, nunca tente interpretar as probabilidades de regressões múltiplas e suas parciais, sem olhar os gráficos parciais.] Em nossa experiência, a fonte mais comum de valores significantes de *P* em regressões múltiplas é a inclusão de um único ponto que foi registrado na escala errada, talvez milímetros ao invés de centímetros, ou a simples erros de digitação. Freqüentemente, estes pontos não se sobressaem em regressões simples, mas ficam óbvios em gráficos de parciais. Veja Anscombe (1973) para exemplos que se aplicam igualmente a regressões simples e gráficos de parciais. Gráficos de parciais são também usados para verificar outras premissas da regressão múltipla, como a de que as relações são lineares e a de que a variabilidade é constante ao longo da linha. Se a inspeção dos gráficos de parciais indicar que alguma coisa está errada e for preciso transformar os dados (veja capítulo 11), ou corrigir ou deletar alguma observação, toda a análise precisará ser refeita, inclusive o exame dos gráficos de parciais. Já que a técnica é baseada na análise dos resíduos em torno de relações, a mudança de uma relação provocará mudanças em todas as outras da análise.

Se as variáveis independentes são correlacionadas (isto é, as variaveis não são independentes), a análise pode identificar muito mais variáveis significantes do que realmente existe (Lennon 2000). Também, nenhuma das averiguações abaixo terá qualquer utilidade, a menos que os dados tenham sido coletados segundo um delineamento amostral adequado.

As análises de variáveis categóricas obedecem aos mesmos princípios, embora muitos livros de estatística tenham seções para variáveis contínuas (regressão múltipla), variáveis categóricas (análise de variância - ANOVA) e mistura de variáveis contínuas e categóricas (análise de covariância - ANCOVA). Todos estes métodos poderiam ser chamados de análise de variância, mas a maioria dos livros não enfatizam as similaridades.

Interações

Assumimos, neste capítulo, que as variações devidas a inúmeras variáveis independentes podem ser adicionadas para se obter a variação global. Entretanto, nem sempre isto é verdade. A figura 50 ilustra uma situação onde os efeitos de duas variáveis independentes cancelam uma a outra.

Neste exemplo, não há um efeito simples de espécies, porque os pesos médios das duas espécies, sem considerarmos o sexo, são os mesmos. Da mesma forma, os pesos médios dos sexos, sem considerarmos espécies, são iguais. Entretanto, dentro de espécies, há um efeito óbvio de sexo, e dentro de cada sexo há um efeito forte de espécies.

Se adicionarmos os efeitos simples de sexo e espécie (ambos são aproximadamente zero) à variabilidade dentro de cada sexo por nível de espécie (o resíduo), não teremos a variabilidade total, que é muito maior. Matematicamente, isto pode ser representado pela seguinte inequação:

$$V_{FATOR1} + V_{FATOR2} + V_{RESÍDUO} < V_{TOTAL}$$

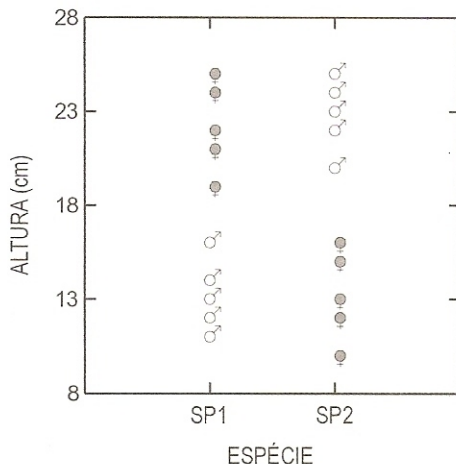
Podemos inventar um fator "fantasma", chamado termo de interação, para equilibrar a equação:

$$V_{FATOR1} + V_{FATOR2} + V_{RESÍDUO} + V_{INTERAÇÃO} = V_{TOTAL}$$

Uma interação indica que o efeito de um (ou mais) fatores depende dos níveis de outros fatores. Interações podem ocorrer tanto em variáveis categóricas quanto em variáveis contínuas e entre variáveis contínuas e categóricas. Elas são mais freqüentemente estudadas com variáveis categóricas, porque a maioria dos programas estatísticos para computadores automaticamente sempre levam em conta todas as interações possíveis para um dado modelo de ANOVA. Entretanto, isto nem sempre é adequado e iremos discutir a seleção de variáveis no próximo capítulo. Em termos de nosso fluxograma conceitual, a presença de interação indica que deixamos de fora uma ou mais variáveis importantes. Em geral, descobrir interações não é um exercício interessante por si próprio. Entretanto, interações são sinais importantes de que precisamos repensar nosso fluxograma de forma que possamos entender porque as interações ocorrem.

FIGURA 50

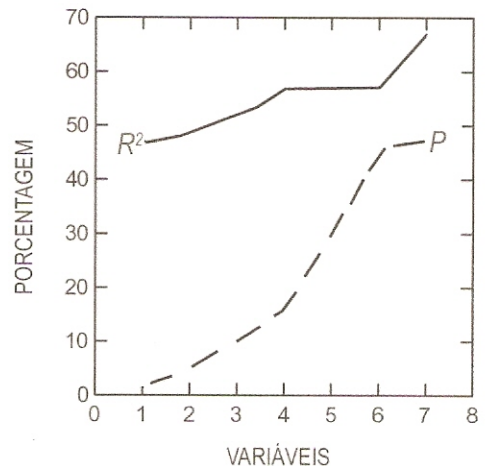
Lembrem-se que o cálculo das razões de *F*, necessárias para se estimar as probabilidades em ANOVA e regressão, envolve médias quadradas, que são variâncias compostas (capítulo 6). Para se determinar quais variâncias estão incluídas em cada média quadrada, temos que levar em conta se os fatores são fixos ou randômicos. Alguns programas de computadores perguntarão se os fatores são fixos ou randômicos, e calcularão as razões de *F* corretas. Outros requerem que o usuário ativamente forneça esta informação e outros só são capazes de calcular razões de *F* para um tipo de fator (em geral, fixo). Portanto, se o leitor fizer os cálculos manualmente, ou usar um programa que não solicita informação sobre se os fatores são fixos ou randômicos, deverá consultar um livro de estatística apropriado. Os leitores com facilidade em matemática podem consultar Winer *et al.* (1991) para aprender como construir as razões de *F* corretas. Zar (1996) traz um apêndice que apresenta como fazer isto sem entrar em detalhes do processo envolvido.

Muitas conclusões na literatura ecológica foram baseadas em análises incorretas, porque os pesquisadores analisaram os dados como se todos os fatores fossem fixos (Bennington & Thayne 1994, Newman *et al.* 1997). Revisem o capítulo 7 para ver como este modelo é muito restritivo. Não vamos nos ater nestes pontos estatísticos porque, se os dados foram coletados segundo um delineamento adequado, qualquer estatístico pode ajudar em sua análise. Entretanto, é importante compreender que modelos mistos de ANOVA serão muito fracos, a não ser que haja um número grande de réplicas e de níveis dos fatores randômicos, de modo que as estimativas de variância devido aos níveis amostrados possam ser adequadamente analisadas. Koele (1982) disse que: "Experimentos que tenham fatores randômicos com apenas dois ou três níveis devem ser considerados tão absurdos como testes *t* entre amostras com duas ou três observações".

[Inclua em suas análises somente os fatores justificados pelo bom senso]

Quais variáveis analisar?

Esta é uma das perguntas mais comuns de estudantes para orientadores, e é uma pergunta que os pesquisadores experientes deveriam se fazer com maior freqüência. A resposta mais simples é incluir todos os fatores que apareceram em seu fluxograma. Entretanto, somente deveriam aparecer nos fluxogramas os fatores cuja inclusão seja altamente justificada pela teoria ou pelo bom senso. Não queremos incluir muitos, especialmente se eles não contribuem com informações a respeito da variável dependente. Lembrem-se de que perdemos um grau de liberdade para cada parâmetro estimado (veja capítulo 5).

Para ilustrar um dos perigos de se incluir variáveis desnecessárias em nossas análises, vamos acompanhar o exemplo seguinte. Calculamos a relação entre densidade de lagostins e um índice de poluição em 12 locais. Uma análise de regressão indicou que a hipótese nula, que era a ausência de relação entre a densidade de lagostins e a poluição era muito improvável ($P = 0,014$). Isto é uma evidência forte de que a poluição afeta os lagostins. A seguir, sorteamos números ao acaso para construir 6 variáveis completamente aleatórias e as incluímos uma a uma na análise. A figura 51 mostra o efeito de se adicionar estas variáveis espúrias na

FIGURA 51

proporção da variação explicada (linha contínua) e a probabilidade de o modelo estar de acordo com a hipótese nula (linha tracejada), ambas expressas em porcentagens.

Depois da adição de três variáveis aleatórias, a regressão não pode ser mais considerada significante à 0,05 (5%). As variáveis adicionais aumentam a proporção da variância aparentemente explicada pela regressão, mas este resultado é espúrio, já que os "fatores" acrescentados não passaram de números sorteados ao acaso, por um gerador de números aleatórios. Este é o padrão geral esperado quando o número de variáveis no modelo é alto em relação ao número de observações independentes (em nosso exemplo, diferentes corpos d'água).

Entretanto, algum leitor pode ser sortudo, e a inclusão de uma variável aleatória pode aumentar a "significância" de sua regressão. Isto representaria embasar sua interpretação de dados biológicos nos azares da matemática. Em geral, a adição de variáveis irrelevantes somente irá comprometer as interpretações biológicas. Uma boa regra é decidir quantas repetições serão coletadas, dividi-las por dez e tentar restringir seu modelo para que não inclua mais do que este número de fatores. De qualquer modo, evite delineamentos de experimentos que resultem em menos de dez graus de liberdade no resíduo (Green 1989).

Harris (1975) sugeriu que a análise de regressão múltipla não é robusta a violações de sua premissa sobre a distribuição normal de suas variáveis, a não ser no caso de haver pelo menos 50 observações a mais do que o número de variáveis independentes ! A filosofia de selecionar variáveis tem muito em comum com a filosofia de comparações múltiplas e um artigo de Tukey (1991) é um bom lugar para se começar a entender os tipos diferentes de questões que podem ser formuladas a respeito do mundo real.

Inteligência artificial

Quando temos inúmeros fatores medidos, os métodos estatísticos não são eficientes em escolher quais deles são relacionadas com os outros. Para entender isto, vamos considerar o tipo de resultado produzido por alguns esquemas automatizados de seleção de variáveis. Para os próximos exemplos, usaremos apenas dados produzidos por computadores através de um gerador de números aleatórios . Portanto, não há relação causal entre as variáveis e as relações entre elas são apenas (e exatamente) aquelas esperadas pelos azares da probabilidade.

Um dos esquemas mais comuns de seleção de variáveis é a regressão "step-wise". Existe um grande número de variações neste procedimento, incluindo aquelas denominadas "step-up" (passo acima), "step-down" (passo abaixo) e "best-subsets regression" (regressão dos melhores subconjuntos). Os praticantes defendem uma ou outra como sendo "melhor" do que as demais, embora na verdade isto faça pouca diferença (Berk 1978). Entretanto, nosso ponto é que as probabilidades associadas com os resultados de qualquer uma delas não podem ser relacionados com nenhuma hipótese nula conhecida (p. ex. Freedman 1983, James and McCulloch 1990).

Elas são pseudoprobabilidades e não devem ser apresentadas como se fossem qualquer coisa além disto. Se a técnica for usada apenas para gerar hipóteses, não há razão para apresentar pseudoprobabilidades, que não têm interpretação lógica. Podemos garantir aos leitores que perguntar a um biólogo competente é uma maneira muito melhor de decidir quais variáveis deveriam ser estudadas. A razão disto é porque deveríamos conhecer alguma coisa a respeito do relacionamento das variáveis independentes antes de incluí-las na análise (veja capítulo 10). Apenas como ilustração, vamos imaginar que os leitores têm um conhecimento detalhado das relações entre as variáveis independentes, mas não podem gerar uma hipótese sobre quais são prováveis de afetar a variável dependente.

A tabela 7 foi gerada por nosso computador usando o pacote estatístico SYSTAT. Nela se vê os resultados de uma análise de regressão múltipla convencional, relacionando a densidade de lagostins com 10 potenciais variávies independentes. Atribuímos a elas nomes realísticos, mas lembrem-se que foram criadas em um gerador de números aleatórios.

TABELA 7

Dep Var: LAGOSTIN N: 14 Multiple R: 0.928 Squared multiple R: 0.860
Adjusted squared multiple R: 0.395 Standard error of estimate: 0.235

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	0.941	0.627	0.000	.	1.500	0.231
VELOCIDADE	-0.131	0.520	-0.108	0.256	-0.253	0.817
TEMPERATURA	-0.122	0.401	-0.109	0.361	-0.304	0.781
PROFUNDIDADE	0.320	0.345	0.289	0.477	0.926	0.423
LARGURA	-0.428	0.328	-0.421	0.448	-1.305	0.283
COBERTURA	0.140	0.482	0.134	0.221	0.291	0.790
GALHOS_ SUBMERSOS	-0.645	0.308	-0.732	0.380	-2.092	0.128
ROCHAS	0.069	0.378	0.071	0.311	0.184	0.866
GARÇAS	-0.057	0.454	-0.056	0.236	-0.126	0.907
PEIXES/CARNIVOROS	-0.212	0.438	-0.177	0.347	-0.484	0.661
PEIXES/HERBÍVOROS	0.359	0.263	0.368	0.640	1.363	0.266

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	1.024	10	0.102	1.848	0.335
Residual	0.166	3	0.055		

Não se preocupe se ainda não entende a maior parte da tabela. A última coluna fornece a probabilidade de cada variável (primeira coluna) estar de acordo com a hipótese nula. Nenhuma delas foi significante ao nível de 0,05. Isto é esperado, já que a análise foi baseada em variáveis aleatórias. Entretanto, não teríamos nos surpreendido se uma ou duas das dez probabilidades tivessem sido menores que 0,05, porque, a este nível de significância, esperamos que cerca de um em vinte resultados apareça como significante quando a hipotese nula não é falsa. A tabela de análise de variância no final de nossa tabela indica que a probabilidade geral de nosso modelo estar de acordo com a hipótese nula é 0,335. Isto é reconfortante, porque sabemos que os dados se conformam exatamente com a hipótese nula.

Usamos a opção "stepwise" para selecionar as "melhores" variáveis a partir do mesmo arquivo de dados e o algorítmo "stepwise" selecionou três variáveis. Alguns programas como SYSTAT, apresentam as pseudo-probabilidades entre aspas, e não fornecem uma "probabilidade" geral para a regressão. Entretanto, muitos programas não são tão cuidadosos e irão apresentar as probabilidades de regressões "stepwise". Neste caso, para o nosso exemplo, teríamos o resultado apresentado na tabela 8 em que as três variáveis selecionadas aparecem como altamente "significantes". Na verdade, a pseudo-probabilidade indica apenas duas chances em 1000 de que "galhos_submersos" se conforme com a hipótese nula e a da regressão geral foi de apenas três em 1000. Pela forma que criamos os dados, sabemos, desde o início, que eles se conformam a hipótese nula convencional (nenhuma relação entre as variáveis independentes e a variável dependente). Portanto, não temos nenhuma idéia de com qual hipótese nula estas pseudo-probabilidades se relacionam.

TABELA 8

Dep Var: LAGOSTIN N: 14 Multiple R: 0.856 Squared multiple R: 0.733
Adjusted squared multiple R: 0.653 Standard error of estimate: 0.178

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	0.951	0.135	0.000	.	7.045	0.000
LARGURA	-0.448	0.166	-0.440	0.996	-2.690	0.023
GALHOS_ SUBMERSOS	-0.592	0.145	-0.672	0.989	-4.095	0.002
PEIXES/HERBÍVOROS	0.406	0.160	0.416	0.993	2.536	0.030

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	0.873	3	0.291	9.165	0.003
Residual	0.317	10	0.032		

Não há razão para apresentar os resultados de regressões "stepwise" em publicações. As hipóteses baseadas na opinião de especialistas serão, em regra, mais úteis, além do que poucos de seus leitores entenderão que as pseudoprobabilidades apresentadas não são probabilidades em qualquer senso convencional. Note que as estatísticas apresentadas na tabela 8 não se relacionam de uma maneira inteligível com os resultados da tabela 7 e que, além disso, o algorítmo "stepwise" pode encontrar outra combinação de variáveis que descreva os dados igualmente "bem". Por exemplo, a tabela 9 mostra outro "melhor" modelo obtido por regressão "stepwise", quando ajustamos apenas uma das opções da análise, alterando-a ligeiramente. Em tudo é a mesma análise, exceto que mudamos o valor de *P* para entrada ou remoção de variáveis no modelo. Na análise da tabela 8 este valor foi de 0,15, o valor "default" de SYSTAT (Wilkinson 1998). Na tabela 9, usamos $P = 0,20$, o valor "default" de alguns outros programas.

O modelo agora incluiu outra variável e as pseudoprobabilidades mudaram drasticamente. Mudanças em outras opções ou procedimentos podem mudar o "melhor" modelo ainda mais dramaticamente. O leitor que desejar deixar o computador pensar por si, deve esperar até reunir dados para testar as hipótese do computador, antes de publicá-las.

Deve ter ficado claro que incluir variáveis espúrias na análise pode ser tão danoso quanto deixar de fora variáveis importantes. **Medir tudo e deixar "os dados falarem por si mesmos" não é uma maneira eficiente de descobrir coisas.** Por esta razão, a maioria das revistas científicas exigem que os pesquisadores testem suas hipóteses. Se cada pesquisador publicasse cada hipótese que gerasse sem ter que testá-las, o volume das publicações aumentaria milhares de vezes, enquanto a comunicação diminuiria na mesma proporção (veja Platt 1964). Hipóteses não testadas só deveriam ser publicadas quando prometem a mudança de paradígmas (Kuhn 1970). Harris (1975) opinou que "A estatística é uma forma de controle social sobre o comportamento profissional dos pesquisadores. A justificativa final de qualquer procedimento estatístico reside no tipo de comportamento que ele encoraja ou desencoraja nos pesquisadores". Deste ponto de vista, os procedimentos "stepwise" geralmente encorajam comportamentos antissociais.

TABELA 9

Dep Var: LAGOSTINS N: 14 Multiple R: 0.888 Squared multiple R: 0.789
Adjusted squared multiple R: 0.695 Standard error of estimate: 0.167

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	1.122	0.168	0.000	.	6.659	0.000
LARGURA	-0.349	0.169	-0.343	0.853	-2.071	0.068
GALHOS SUBMERSOS	-0.628	0.138	-0.713	0.962	-1.537	0.001
PEIXES/CARNÍVOROS	-0.339	0.221	-0.284	0.689	-1.537	0.159
PEIXES/HERBÍVOROS	0.294	0.167	0.301	0.804	1764	0.111

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	0.939	4	0.235	8.402	0.004
Residual	0.251	9	0.028		

Variáveis fantasmas geradas por computadores

Programas de computadores podem nos enganar fazendo-nos pensar que não precisamos usar de bom senso quando estivermos decidindo sobre quais variáveis devemos estudar. Além disso, eles podem criar variáveis espúrias. Os modelos que consideramos até agora não investigam interações (veja o capítulo anterior se esqueceu o que é uma interação). Entretanto, muitos programas que executam análise de variância podem gerar todas as possíveis interações automaticamente, quer o pesquisador queira ou não. Considere os resultados de um pesquisador que executou um experimento controlado para testar o efeito de cinco fatores sobre a densidade de lagostins, usando dois níveis por fator. O pesquisador, um estudante de pós-graduação, fica chocado ao descobrir que nenhum dos fatores sugeridos por seu orientador foi significante (tabela 10).

TABELA 10

Dep Var: LAGOSTIN N: 96 Multiple R: 0.196 Squared multiple R: 0.038

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
PROFUNDIDADE	0.130	1	0.130	1.596	0.210
GALHOS/SUBMERSOS	0.120	1	0.120	1.468	0.229
VELOCIDADE	0.010	1	0.010	0.119	0.731
PEIXES/HERBÍVOROS	0.032	1	0.032	0.396	0.531
PEIXES/CARNÍVOROS	0.000	1	0.000	0.001	0.974
ERROR	7.354	90	0.082		

Então, ele decide testar todas as interações possíveis (tabela 11). **Tabelas de ANOVA devem ser interpretadas de baixo para cima.** Se o fator está envolvido em uma interação significante, então todos os termos acima, nos quais o fator está envolvido, são presumivelmente significantes. No entanto, os pesquisadores algumas vezes concluem que o efeito simples de um fator envolvido em uma interação não é significante. Isto é ilógico. Se o fator está envolvido em uma interação, então não existe resposta simples. O efeito do fator depende do nível do outro fator ou fatores envolvidos na interação. Baseado na tabela 11, o estudante concluiu que há uma interação significante ($P = 0{,}001$) entre galhos/submersos e peixes/herbívoros, e portanto, que galhos/submersos e peixes/herbívoros afetam a densidade de lagostins.

Há muitas explicações biológicas possíveis para uma interação entre galhos_submersos e peixes_herbívoros e isto permite uma seção extensa na discussão da tese dedicada a este ponto. O estudante considera ainda a probabilidade da hipótese nula de que não há interação entre profundidade, velocidade e peixes_herbívoros tão baixa ($P = 0{,}075$) que indique uma tendência que vale a pena ser discutida. O estudante e seu orientador ficam contentes, por que agora há uma tese volumosa para ser defendida, mas a ciência terá avançado?

Este exemplo também foi baseado em dados sorteados ao acaso. Há 26 possíveis termos de interação em uma ANOVA com cinco fatores. Portanto, esperamos em média um resultado "significante" ao nível de 0,05 para dados ao acaso. Entretanto, as interações envolvem mais do que um fator, e em conseqüência esperamos que mais de um fator apareça como "significante". O estudante tinha quase garantia de encontrar resultados "significantes", mesmo que suas variáveis não tivessem nenhuma relação entre sí, como nos dados randômicos do exemplo.

Muitos autores parecem não enxergar que programas de ANOVA não controlam a taxa de erro geral (e.g. Fowler 1990). Muitos tipos de tabelas estatísticas apresentam grande número de testes e também não controlam a taxa de erro geral (Rice 1989). Harris (1975) salienta que programas de regressão múltipla usualmente apresentam um teste de significância da regressão geral, mas que programas de ANOVA para dados categóricos não fazem isso. Isto é estranho, já que a ocorrência de um resultado geral não significante coexistindo com efeitos parciais significativos é um bom indicador de que incluímos variáveis demais em nosso modelo e que a "significância" estatística das parciais podem não passar de artefatos, seja em regressão ou em ANOVA.

É importante identificar interações quando elas ocorrem em nossos modelos, principalmente porque indicam que deveríamos estar incluindo outros fatores, ou reconsiderando as escalas nas quais as variáveis foram medidas. Entretanto, se permitirmos que o computador gere todas as possíveis interações, sem refletirmos quais devem ser as mais apropriadas, estaremos apenas gerando muitas relações espúrias e muita confusão. Alguns pesquisadores, com mais habilidade em matemática, poderão usá-las para ajudar a decidir entre diferentes modelos candidatos (p. ex. Burnham and Anderson 1998). Entretanto, o pesquisador nunca pode abrir mão da responsabilidade de usar a lógica para decidir quais as variáveis devem ser incluídas. Fazer isto bem feito é a arte do cientista.

Algumas vezes desejamos coletar dados para construir hipóteses e dados para testar hipóteses ao mesmo tempo. Neste caso, precisamos de grande número de dados para poder dividí-los em um sub-grupo de dados exploratórios e outro de dados de validação. Os dados exploratórios usualmente resultam em estatísticas superotimistas em relação à capacidade de previsão do modelo (Picard and Cook 1984). Se o sub-grupo de dados de validação é representativo da amostra exploratória, e se ambos forem representativos da população de interesse, este processo pode economizar tempo e dinheiro. A ciência trabalha basicamente testando o quanto os resultados podem ser repetidos. Sub-grupos de validação não são evidência tão boa quanto uma repetição mais substancial, feita por outro pesquisador, em outro local e tempo (Guttman 1985). Ainda assim, muitas vezes é o melhor que podemos fazer.

O problema geral em não publicar resultados não significativos se extende a comparações entre estudos, interpretação da literatura, revisões e técnicas de meta-análise (Palmer 1999, Thornhill et al 1999). Em todos estes níveis, o pesquisador deve se perguntar o que foi testado, sem ter sido relatado, mas não vamos mais nos ater a essa questão.

TABELA 11

Dep Var: LAGOSTIN N: 96 Multiple R: 0.629 Squared multiple R: 0.395

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
PROFUNDIDADE	0.130	1	0.130	1.804	0.184
GALHOS_SUBMERSOS	0.120	1	0.120	1.659	0.202
VELOCIDADE	0.010	1	0.010	0.135	0.715
PEIXES_HERBÍVOROS	0.032	1	0.032	0.448	0.506
PEIXES_CARNÍVOROS	0.000	1	0.000	0.001	0.973
PROFUNDIDADE* GALHOS_SUBMERSOS	0.067	1	0.067	0.923	0.340
PROFUNDIDADE*VELOCIDADE	0.092	1	0.092	1.267	0.265
PROFUNDIDADE*PEIXE_HERBÍVOROS	0.231	1	0.231	3.199	0.078
PROFUNDIDADE*PEIXE_CARNÍVOROS	0.025	1	0.025	0.341	0.561
GALHOS_SUBMERSOS*VELOCIDADE	0.001	1	0.001	0.015	0.903
GALHOS_SUBMERSOS*PEIXES_HERBÍVOROS	0.956	1	0.956	13.223	0.001
GALHOS_SUBMERSOS*PEIXES_CARNÍVOROS	0.046	1	0.046	0.643	0.426
VELOCIDADE*PEIXES_HERBÍVOROS	0.023	1	0.023	0.321	0.573
VELOCIDADE*PEIXES_CARNÍVOROS	0.036	1	0.036	0.503	0.481
PEIXES HERBÍVOROS*P.CARNÍVOROS	0.017	1	0.017	0.240	0.626
PROFUNDIDADE*GALHOS_SUBMERSOS* VELOCIDADE	0.087	1	0.087	1.202	0.277
PROFUNDIDADE*GALHOS_SUBMERSOS* P.HERBÍVOROS	0.022	1	0.022	0.309	0.580
PROFUNDIDADE*GALHOS_SUBMERSOS* P.CARNÍVOROS	0.123	1	0.123	1.700	0.197
PROFUNDIDADE*VELOC.*P.HERBÍVOROS	0.231	1	0.231	3.194	0.075
PROFUNDIDADE*VELOC.*P.CARNÍVOROS	0.129	1	0.129	1.785	0.186
PROFUNDIDADE*P.HERBÍVOROS* P.CARNÍVOROS	0.003	1	0.003	0.040	0.843
GALHOS_SUBMERSOS*VELOCIDADE* P.HERBÍVOROS	0.051	1	0.051	0.710	0.403
GALHOS_SUBMERSOS*VELOCIDADE* P.CARNÍVOROS	0.103	1	0.103	1.424	0.237
GALHOS_SUBMERSOS*P.HERBÍVOROS* P.CARNÍVOROS	0.007	1	0.007	0.093	0.761
VELOC.*P.HERBÍVOROS*P.CARNÍVOROS	0.099	1	0.099	1.366	0.247
PROFUNDIDADE*GALHOS_SUBMERSOS* VELOCIDADE*P.HERBÍVOROS	0.006	1	0.006	0.088	0.768
PROFUNDIDADE*GALHOS_SUBMERSOS* VELOCIDADE*P.CARNÍVOROS	0.000	1	0.000	0.000	0.984
PROFUNDIDADE*GALHOS_SUBMERSOS* P.HERBÍVOROS*P.CARNÍVOROS	0.113	1	0.113	1.557	0.217
PROFUNDIDADE*VELOCIDADE* P.HERBÍVOROS*P.CARNÍVOROS	0.014	1	0.014	0.187	0.667
GALHOS_SUBMERSOS*VELOCIDADE* P.HERBÍVOROS*P.CARNÍVOROS	0.115	1	0.115	1.597	0.211
PROFUNDIDADE*GALHOS_SUBMERSOS* VELOC.*P.HERBÍVOROS*P.CARNÍVOROS	0.131	1	0.131	1.817	0.200

CAPÍTULO 10

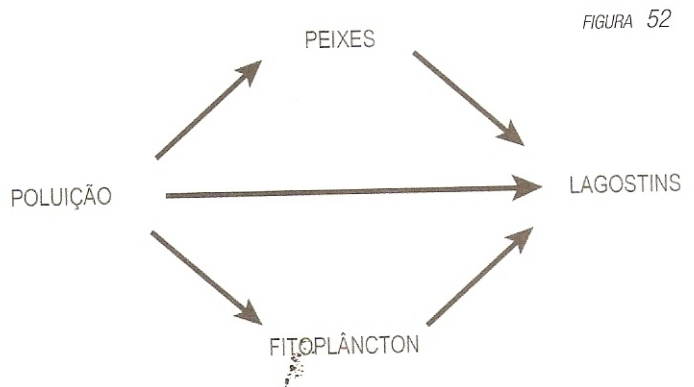
> Na ausência de fluxogramas, as análises estatísticas não podem ser interpretadas

Modelos complexos

Até agora temos considerado como analisar situações muito simples, nas quais as variáveis independentes podem afetar a variável dependente, mas não podem afetar umas às outras. É claro que isto não é muito realista. Neste capítulo, vamos continuar a lidar com relações lineares simples, mas vamos permitir que as variáveis independentes se afetem mutuamente. Existem outras maneiras de se atacar problemas semelhantes, além da que vamos apresentar aqui, mas geralmente as interpretações são mais difíceis (p. ex. Mac Nally 2002).

Vamos analisar o exemplo dado no capítulo 2. O fluxograma apresentado na figura 52 mostra a direção da influência, mas não especificamos o que flui ao longo das setas.

FIGURA 52

Na tabela 12, apresentamos os dados referentes a 30 lagoas onde as quatro variáveis foram medidas simultaneamente. "Poluição" representa a concentração de metal pesado em partes por bilhão, "peixes" é medido como o número médio de peixes capturados por hora por rede de espera, "fitoplâncton" é a concentração de clorofila relativa a um padrão, e "lagostins" é o número de lagostins capturados por armadilha por hora. Quando tratamos de relações simples, diferenças nas escalas de medidas não importavam muito. Poderíamos dizer coisas como "o aumento de uma unidade de poluição levará à diminuição de tantas unidades de

lagostins". Entretanto, se os efeitos de poluição em fitoplâncton são medidos como unidades de metal pesado, e os efeitos de fitoplâncton em lagostins são medidos em termos de clorofila, o mesmo efeito não pode fluir ao longo de ambas as setas de nosso fluxograma.

Para colocar todas as variáveis na mesma escala, podemos dividir o valor de cada variável pelo desvio padrão desta variável, como Student fez para remover os problemas de escala de sua estatística *t*. Os dados originais e os desvios padrões têm a mesma dimensão, portanto, quando dividimos uma pela outra, temos uma quantidade aparentemente adimensional (as unidades do numerador e do denominador se cancelam).

TABELA 12

POLUIÇÃO	PEIXES	FITOPLÂNCTON	LAGOSTINS
9,5	9,0	12,5	2,4
8,5	9,9	11,6	5,0
9,2	9,1	15,3	5,8
10,3	8,0	13,9	6,2
9,6	9,1	14,4	3,6
10,9	5,6	14,2	5,2
10,7	4,3	15,1	8,1
8,9	6,2	12,0	3,4
10,5	6,2	14,2	4,4
9,0	7,9	14,3	4,0
9,8	8,0	15,4	4,7
8,0	8,0	11,1	3,4
9,0	8,7	12,7	4,0
9,8	5,4	12,9	6,9
10,8	6,1	15,8	6,4
9,2	8,0	13,3	3,4
10,4	7,5	14,2	4,0
9,4	7,5	16,0	9,1
8,6	10,2	14,2	6,8
9,8	6,8	13,3	6,5
8,7	8,3	12,0	5,8
8,8	7,9	12,5	4,4
9,8	7,8	15,3	6,1
8,2	9,5	13,0	3,5
9,9	6,6	15,7	9,3
8,7	7,0	14,3	5,1
8,9	7,7	13,0	3,0
10,0	8,3	15,1	4,8
8,1	9,8	14,4	3,0
10,2	7,3	15,7	6,3

Não se importem se não entendem a álgebra. Efetivamente, ao dividirmos pelo desvio padrão, estamos colocando todas as medidas em unidades de desvio padrão. Agora, podemos dizer coisas como "Um aumento de um desvio padrão em poluição levará ao decréscimo de tantos desvios padrões no número de lagostins". Quando calculamos estatísticas baseadas em dados padronizados, elas são chamadas estimativas padronizadas de parâmetros. Muitos programas de computadores fornecem as estimativas de parâmetros padronizadas mas, em geral, elas são de pouco uso, a menos que estejamos interessados em ver como os efeitos se propagam através dos fluxogramas.

O uso de coeficientes padronizados para avaliar cadeias de efeitos é chamado de "Análise de Caminhos" (*"Path Analysis"*) ou *"Modelagem de Equação Estrutural". Vamos mostrar como fazer isto de uma forma simples, de forma que possam* entender os conceitos e remeter-se à literatura, onde poderão encontrar detalhes sobre os métodos mais aceitos e empregados. Antes que se envolvam em demasia com os aspectos matemáticos do método, gostaríamos que entendessem alguns conceitos relacionados com efeitos diretos, indiretos e efeitos gerais, sem o que, é impossível desenhar esquemas amostrais efetivos.

Estimando efeitos diretos

Primeiramente, vamos analisar os dados da tabela 12 usando uma regressão múltipla, mas utilizando os coeficientes padronizados da regressão. Como vimos, eles são adimensionais e podemos usá-los para avaliar a contribuição relativa das variáveis independentes, para a variação observada na variável dependente. A equação resultante foi:

$$\text{Lagostins} = 0{,}0 - 0{,}16 * \text{Poluição} - 0{,}39 * \text{Peixe} + 0{,}55 * \text{Fitoplâncton}$$

Podemos representar isto no fluxograma, atribuindo valores para os coeficientes padronizados para representar a força de cada relação (figura 53). Neste contexto, estes valores são chamados de coeficientes de "caminhos" (*"path"*).

Baseado na regressão múltipla, o efeito da poluição é negativo e sua magnitude (0,16) é menor do que a do efeito dos peixes (0,39) e do fitoplâncton (0,55). Estatisticamente, não parece haver um efeito significante de poluição, já que a regressão múltipla estimou uma probabilidade alta ($P = 0{,}53$) para esta parcial, mas há alguma evidência de um efeito de peixes ($P = 0{,}07$) e uma forte indicação de um efeito de fitoplâncton ($P = 0{,}01$). Isto vai contra a intuição, porque geralmente há mais lagostins em lagoas mais poluídas e uma regressão simples indicou um efeito positivo significante ($P = 0{,}03$) de poluição sobre os lagostins, com uma magnitude de 0,41.

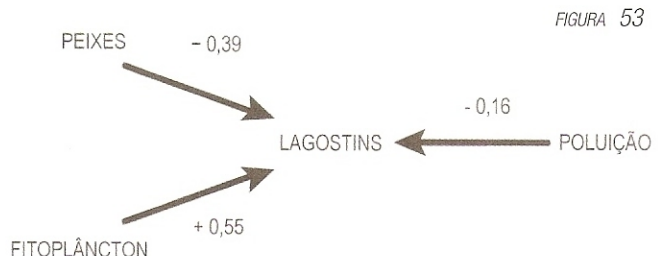



FIGURA 53

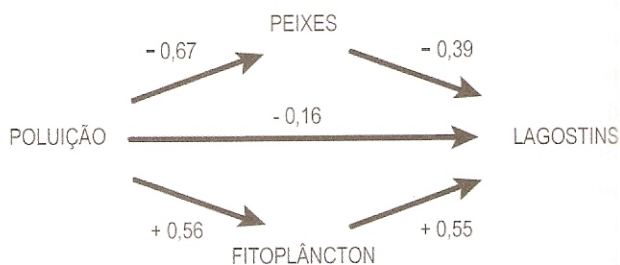
Regressões simples podem ser enganosas (capítulo 8). Entretanto, podemos ver que o fluxograma representado pela regressão múltipla (figura 53) não representa o sistema como acreditamos que ele funcione (figura 52). A figura 53 tem a forma de uma estrela, com todas as setas dirigindo-se para a variável dependente, em todos os ângulos. As direções das setas são arbitrárias, porque as variáveis independentes não afetam umas às outras. Sistemas que podem ser representados por um fluxograma em forma de estrela são muito raros em ecologia, mas a maioria dos procedimentos estatísticos comumente empregados assumem este formato. É por este motivo que devemos construir um fluxograma que represente a maneira como acreditamos que o sistema em estudo funcione, antes de escolher a análise.

A regressão múltipla está nos dizendo que se peixes e fitoplâncton forem mantidos constantes, o efeito da poluição sobre os lagostins será negativo e, possivelmente, não significante. O problema é que, no mundo real, não é possível manter as densidades de peixes e fitoplâncton constantes, quando este tipo de poluição varia. Talvez, se extinguíssemos todos os peixes e fitoplâncton dos lagos poderíamos ver o efeito previsto pela regressão múltipla, mas espero que concordem que isto seria eticamente questionável, além de difícil em sua execução. O conhecimento de efeitos diretos pode ser útil em campos como o da medicina ou agricultura, onde podemos manipular alguns fatores além dos níveis que eles ocorrem naturalmente e eliminar ou controlar outros quase totalmente. Em situações em que não possuímos este controle, será muito mais útil conhecer os efeitos reais de um fator, e não seus efeitos hipotéticos, se tudo o mais fosse mantido constante.

Estimando efeitos indiretos

Podemos usar a análise de caminhos para investigar tanto os efeitos diretos quanto os indiretos. Os coeficientes de regressão padronizados nos revelam os efeitos diretos hipotéticos das variáveis que têm setas diretamente ligadas aos lagostins (figura 54).

Notem que os efeitos diretos das variáveis que afetam diretamente os lagostins são os mesmos da figura 53. Mas agora aparecem também os efeitos diretos da poluição sobre os peixes e da poluição sobre o fitoplâncton. Podemos obter as regressões padronizadas para estas conexões por regressões simples. Para calcular os efeitos indiretos, multiplicamos os coeficientes de "caminhos" ao longo de cada fluxo. Para obter o efeito geral de uma variável sobre outra, somamos os efeitos diretos e indiretos.

FIGURA 54

Para poluição, temos os seguintes fluxos e seus coeficientes de "caminhos" correspondentes:

Poluição — Lagostins	–0,16
Poluição — Peixes — Lagostins	+0,26
Poluição — Fitoplâncton — Lagostins	+0,31

O efeito geral (a soma dos efeitos diretos e indiretos) da poluição é 0,41, que coincide com o resultado que obtivemos pela regressão simples de lagostins e poluição. Embora a regressão simples tenha fornecido uma resposta numérica correta, ela indicou um efeito direto positivo de poluição sobre lagostins que não existe. A regressão múltipla forneceu a resposta correta em relação aos efeitos diretos, mas isto é uma resposta a respeito de um mundo que não acreditamos existir. Em nosso modelo, poluição está dirigindo o sistema, embora só tenha um pequeno efeito direto sobre lagostins.

Um aspecto perturbador das análises convencionais é que elas tendem a desconsiderar os efeitos finais dos fatores, focalizando-se nas causas próximas, que na verdade são apenas o resultado das variáveis mais importantes da porção mais a esquerda do modelo. Experimentos manipulativos, onde o pesquisador mantém algumas variáveis constantes ou produzem combinações de níveis de fatores que não existem na natureza, têm a mesma limitação. Eles estão se reportando a efeitos diretos esperados em um mundo imaginário. A análise de caminhos permite-nos interpretar um mundo carregado de sentido biológico. Na ausência de fluxogramas, as análises estatísticas geralmente não podem ser interpretadas.

Algumas pedras na análise de caminhos

Fizemos a análise de caminhos parecer muito atraente. Entretanto, ela tem algumas limitações sérias (Petraitis et al. 1996, Shipley 1999). Usamos esta análise apenas para ilustrar o primeiro passo em direção a lidar com modelos ecológicos mais realistas. As análises de "caminhos" (na verdade, há mais de um tipo) geralmente assumem relações lineares, sem retroalimentação nos fluxos, e são baseadas em desvios padrões. Portanto, se não amostrarmos a variabilidade real de cada variável do sistema, os coeficientes de "caminhos" não vão fornecer os efeitos esperados. Por esta razão, é difícil interpretar a análise de caminhos quando os pesquisadores manipularam experimentalmente algumas variáveis (Petraitis et al. 1996).

Além disso, o uso de coeficientes padronizados de regressões simples pode não ser correto, a não ser que a análise inclua todos os fluxos possíveis. Por exemplo, não incluímos o fluxo de peixes para fitoplâncton em nosso exemplo, porque não tínhamos evidência de uma relação causal. Se não incluirmos todos os fluxos possíveis, Petraitis et al.(1966) sugeriram que temos que "enganar" o computador modificando a matriz que ele usa para calcular as regressões múltiplas. Variáveis categóricas só podem ser colocadas no modelo se elas tiverem apenas dois níveis, ou se pelo menos forem ordinais. Entretanto, a interpretação de uma variável cuja distribuição não pode ser descrita por seu desvio padrão é difícil.

Algumas vezes, a análise de caminhos pode não ser boa para revelar padrões que podem ser mostrados por outros métodos, embora algumas formas de análise de caminhos provavelmente serão sempre necessárias ao se analisar sistemas com escalas múltiplas (O'Neill e King 1998, trazem um bom apanhado sobre problemas de escalas). Portanto, embora sugerimos que todas as análises estatísticas requerem o exame prévio de um fluxograma, o leitor só deveria considerar aplicar modelos de equações estruturais quando estiver lidando com sistemas relativamente simples. Precisará também dispor de um bom programa estatístico e de um estatístico para livrá-lo da confusão matemática em que quase certamente mergulhará.

Neste livro, como na maioria dos livros, nos concentramos em técnicas de análise de variâncias, porque este é um ponto de partida e os conceitos são simples. Entretanto, Guttman (1985) comentou a este respeito "Não é diferente de dar um martelo de brinquedo a uma criança: ela vai usá-lo em tudo que ver". A dificuldade de se estimar parâmetros é relacionada à classe geral de problemas chamada "problemas inversos" (Wood 1997). Se decidirmos analisar o fluxograma matematicamente, temos de considerar a possibilidade de usar simulações intensivas em computadores (p. ex. Starfield e Bleloch 1991) ou outras técnicas matemáticas complexas (p. ex. Link 1999, Spitz e Leks 1999, Wardle 1999, Burnham e Anderson 1998, McCune e Grace 2002). Todas estas técnicas estão muito além do escopo desta introdução e requerem a ajuda de um estatístico competente.

Esperamos que tenhamos ao menos dado uma idéia de porque alguns pesquisadores estão lançando mão de técnicas muito diferentes das técnicas estatísticas padrões, que se tornaram o "emblema" de cientístcas iniciados. No capítulo 12, descreveremos alguns aspectos de análises multivariadas, que à primeira vista são um pesadelo para biólogos e um paraíso para matemáticos, e tentaremos tratá-las sob a mesma abordagem que adotamos no resto deste livro. Entretanto, no próximo capítulo retornaremos a métodos lineares de mínimos quadrados e à análise de relações simples, embora os métodos envolvidos nem sempre sejam tão simples.

[Métodos baseados em mínimos quadrados são apropriados somente sob condições muito restritas]

Endireitando o mundo com transformações e outros truques

Até agora, nossas análises se basearam em modelos lineares nos quais temos minimizado os desvios quadrados para estimar parâmetros. Estes métodos algébricos de mínimos quadrados são teoricamente apropriados somente sob um conjunto de condições muito restritivas. Eles assumem que as relações são lineares, que os efeitos dos fatores são aditivos, que os resíduos dos modelos tenham distribuição normal, que não há erro estocástico na medida das variáveis independentes, que a variação da variável dependente é homogeneamente distribuída ao longo dos níveis da variável independente e ainda outras condições improváveis.

Quando estas condições não são satisfeitas, algum outro modelo será um melhor estimador dos parâmetros. Entretanto, algumas vezes é possível transformar os dados e fazê-los se conformarem às premissas dos mínimos quadrados. Neste capítulo vamos primeiro considerar sobre maneiras de operar estas transformações e depois examinaremos modelos alternativos. Nossos modelos serão muito simples e não nos ateremos em detalhes. Procuraremos esclarecer os aspectos conceituais mais importantes, de forma que elas possam ser usadas para esclarecer e não para obscurecer os padrões da natureza.

Nosso exemplo diz respeito a pessoas tentanto estimar massa ou volume à partir de medidas lineares, digamos, um ecólogo interessado em estimar a biomassa de árvores em uma grande área de floresta. Se preferirem um problema aplicado, imaginem um engenheiro florestal empenhado em estimar o volume de madeira. Biólogos de pesca usam técnicas similares para estimar o peso a partir de medidas de comprimento de peixes, e fisiólogos têm problemas similares quando a massa corpórea afeta as variáveis em que estão interessados. Na realidade, este tipo de relação é esperado em muitos sistemas biológicos (Carlson e Doyle 1999). É relativamente simples tomar medidas do diâmetro de árvores, enquanto a medição da biomassa é tarefa árdua e destrutiva. Portanto, nosso ecólogo decide determinar a relação entre o diâmetro e a biomassa de

30 árvores e usar esta relação para predizer a biomassa florestal à partir de medidas do diâmetro de árvores tomadas sobre toda a área de interesse. Os dados para as 30 árvores estão mostrados na figura 55.

FIGURA 55

A relação não é linear e a variabilidade em biomassa aumenta a medida que o diâmetro aumenta. Não é possível produzir uma equação preditiva útil usando a álgebra de mínimos quadrados diretamente com estes dados. Entretanto, provavelmente a relação conforma uma função potência da seguinte forma:

$$\text{Biomassa} = a * \text{Diâmetro}^b + e_1$$

onde "a" e "b" são coeficientes que descrevem o formato e a posição da curva, enquanto e_1 representa ou a variação aleatória ou variação explicada por fatores que não entraram no modelo e é chamado de "erro". Logaritmizando os dois lados da equação (desconsiderando o "erro"), obtemos um equação linear em uma forma que pode ser tratada pela álgebra convencional de mínimos quadrados.

$$\log(\text{Biomassa}) = \log(a) + b * \log(\text{Diâmetro}) + e_2$$

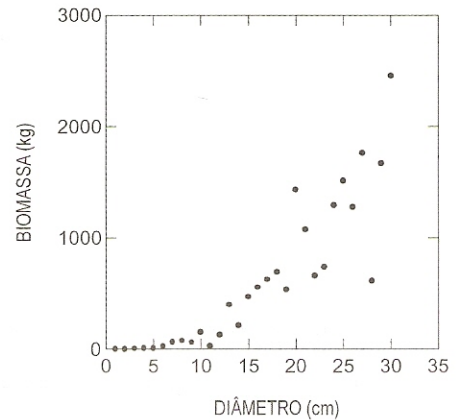
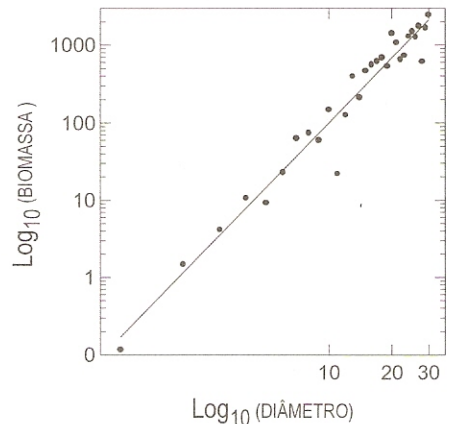
Portanto, o biólogo decide transformar seus dados, colocando ambos os eixos em escalas logarítmicas de base 10 (Figura 56).

FIGURA 56

A relação agora é linear e a variabilidade da biomassa é similar através de todo a amplitude de valores de diâmetros. Isto facilita a matemática e permite ao biólogo usar a álgebra de mínimos quadrados. Vamos deixar o problema de como lidar com o termo de "erro" da equação original para os estatísticos.

A reta da figura 56 pode ser descrita pela equação:

$$\log_{10}(\text{biomassa}) = -0{,}775 + 2{,}778 * \log_{10}(\text{diâmetro})$$

Esta equação preditiva pode ser usada para estimar a biomassa de árvores para as quais apenas o diâmetro foi medido. Tudo parece bem, porque o r^2 (um indicativo grosseiro da capacidade de predição da equação) é alto (0,96) e nós estimamos o coeficiente "b", e o logarítmo do coeficiente "a". Agora podemos tirar o antilogaritmo de –0,775 ($10^{-0,775} = 0{,}168$) para estimar "a" e re-escrever a equação original como:

$$\text{Biomassa} = 0,168 * \text{Diâmetro}^{2,778}$$

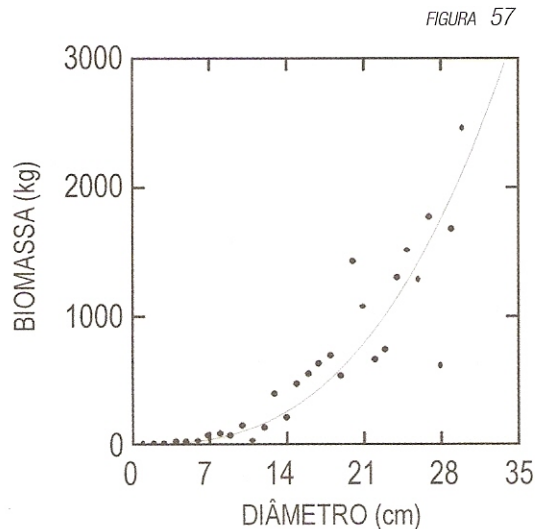
A figura 57 mostra a posição estimada da curva em relação aos dados originais. Podemos ver que aquele r^2 não faz sentido em termos dos dados originais. A linha não explica cerca de 96% da variação nos dados e a variabilidade muda ao longo da linha. A equação é muito boa para estimar a biomassa de árvores pequenas, mas árvores grandes, na faixa de 25-30 centímetros de diâmetro poderiam ter biomassa variando de menos de 1000kg até mais de 2000kg.

[Os pesquisadores as vezes apresentam o gráfico dos dados transformados, mas o ajuste do modelo só pode ser avaliado em relação aos dados originais, não transformados.]

Estimativas por tentativa e erro sem transformação

A grande variabilidade na biomassa das árvores maiores é real e nenhum modelo pode fazê-la desaparecer. Entretanto, a transformação logarítmica reduziu a variabilidade das árvores grandes quando a posição da linha foi estimada. Isto deu tanto peso para árvores pequenas quanto para árvores grandes no momento da estimativa da posição da linha. Isto não era o que o biólogo queria. Pequenos erros na estimativa da biomassa de árvores pequenas terão pequeno efeito na estimativa da biomassa de toda a floresta. Entretanto, mesmo uma mudança diminuta na posição da linha pode mudar a estimativa de biomassa de uma única árvore grande em centenas de quilogramas. Por causa disto, nosso biólogo procura agora por um método que possa preservar a grande variabilidade das árvores grandes e assim dar maior peso a elas na análise. Modelos lineares não são apropriados, portanto ele usa um computador para aplicar uma técnica intensiva de estimativa não-linear.

FIGURA 57

"Modelos não-lineares" são parcialmente tentativa e erro e parcialmente matemáticos. O programa começa com estimativas arbitrárias dos parâmetros, aplica o modelo (o que é equivalente a desenhar a linha no gráfico) e calcula os desvios dos pontos até a linha. Então ele tenta novamente com valores um pouco diferentes, e assim sucessivamente, até conseguir um modelo que resulte nos menores desvios dos pontos observados até a linha. A

matemática está envolvida porque o pesquisador precisa informar a forma geral da equação para o computador, e porque o programa usa algorítmos matemáticos para se assegurar que está mudando as estimativas na direção que vai resultar nos menores desvios. Estes métodos são muito efetivos. O computador fez 16 tentativas (chamadas iterações) para estimar a seguinte equação:

$$\text{Biomassa} = 1{,}139 * \text{Diâmetro}^{2{,}178}$$

Os valores são ligeiramente diferentes daqueles estimados por transformação dos dados, mas a linha resultante ocupa uma posição similar. Na figura 58 o modelo estimado por iterações foi representado pela linha contínua, enquanto a linha tracejada representa a mesma relação representada na figura 57, obtida por transformação logarítmica.

A vantagem para o biólogo é que o método não linear ajustou melhor a linha aos dados das árvores grandes, ao custo de superestimar a maioria das árvores pequenas. Um modelo deste tipo seria melhor para estimar a biomassa da floresta, mas pode ser inapropriado para questões fisiológicas associadas com as árvores pequenas.

Outros métodos atípicos

Modelos não-lineares também são úteis para estimar parâmetros quando não queremos usar mínimos quadrados. As técnicas padrões de mínimos quadrados geralmente são as melhores para predição, mas podem não ser as melhores para descrever as relações de maneira acurada. Considere a relação entre comprimento e altura do corpo para uma espécie de peixe. Para dados morfométricos, a regressão do maior eixo reduzido (média geométrica) pode ser o modelo mais apropriado (p. ex. Ricker 1973).

É fácil de se estimar esta regressão, instruindo-se o programa a usar uma outra equação (chamada função de perda ou "*loss function*") para minimizar a soma das áreas dos triângulos formados entre os pontos e a linha do gráfico (veja capítulo 7), ao invés de minimizar a soma das áreas dos quadrados, como no caso das análises regulares de mínimos quadrados. A figura 59 mostra a posição da linha quando minimizamos os desvios quadrados (linha contínua) e desvios segundo o modelo do maior eixo reduzido (linha tracejada).

A diferença não foi muito grande neste caso, embora possa ser importante em alguns casos, especialmente se o pesquisador estiver analisando os dados em seções, ao invés de com toda a amplitude de variação da variável independente (Ricker 1973). O ponto é que qualquer resíduo pode ser minimizado por iteração, mesmo se não houver uma fórmula matemática simples para fazê-lo.

FIGURA 58

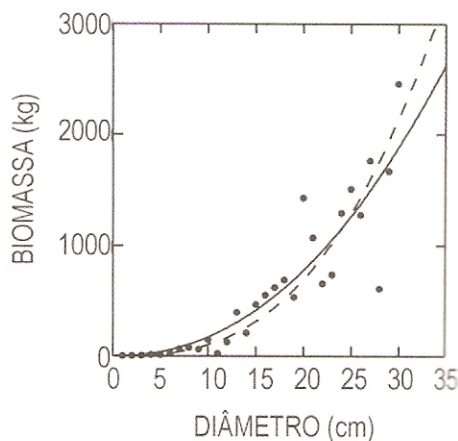
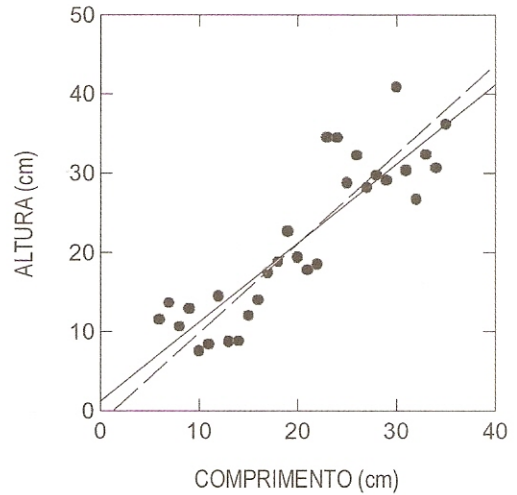



FIGURA 59

Modelos gerais lineares

Em alguns casos, as premissas das técnicas de mínimos quadrados são seriamente violadas. O caso mais comum é quando a variável dependente é medida em uma escala binária, do tipo 0 ou 1 (ou presença/ausência, efeito/ausência-de-efeito, morte/sobrevivência, etc.). Neste caso, os pesquisadores usam regressão logística para analisar os dados. A maioria dos métodos que discutimos neste capítulo se enquadra dentro da técnica geral chamada de modelos gerais lineares. Um modelo é especificado e, se necessário, os dados são transformados de modo a serem linearizados.

Na modelagem geral linear, o modelo, no qual as transformações se baseiam, é chamado de "função de ligação". Para linearizar a função potência, aplicamos a mesma transformação nos dois lados da equação, de modo a manter a igualdade. Entretanto, nem todas as transformações para se obter a função de ligação terão esta propriedade e a escolha de que tipo de transformação empregar e de qual desvio será minimizado (veja figura 41, capítulo 7 para exemplos de distâncias que podem ser minimizadas) depende da questão que se quer responder, da forma esperada da distribuição dos resíduos, e na forma como os dados foram coletados. Em alguns casos, pode ser possível fazer isto algebricamente, como em mínimos quadrados, mas geralmente uma técnica iterativa é usada.

[Mesmo quando as relações não podem ser linearizadas, estimativas razoáveis dos parâmetros podem ser obtidas por técnicas de tentativa e erro (iteração).]

Verossimilhança máxima

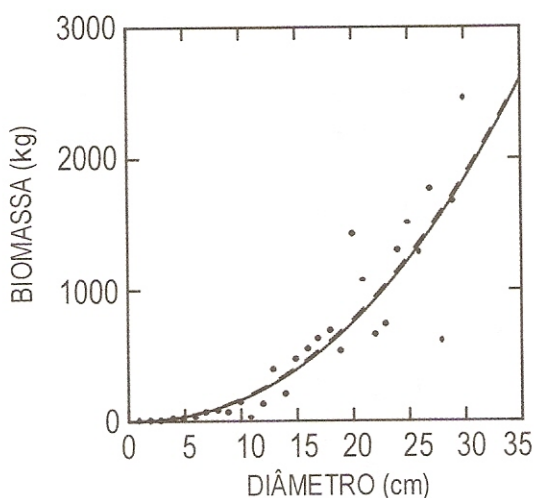
As técnicas não-lineares, como a de verossimilhança máxima, não precisam minimizar resíduos. Esta técnica estima os parâmetros escolhendo os valores que maximizam a probabilidade de se encontrar justamente os valores observados. Sokal e Rohlf (1995) opinaram que "a abordagem da verossimilhança máxima para ajustar uma linha de regressão para dados bivariados ou multivariados é o método mais geral e correto". Entretanto, para amostras pequenas, o método de verossimilhança

máxima não é resistente a erros sistemáticos (viéses) e nem eficiente. Para uma distribuição normal, com covariância conhecida, a verossimilhança máxima se reduz à mínimos quadrados ponderados, sendo os pesos dados pelo inverso da matriz de covariância (Bard 1974). Os leitores com mentes matemáticas podem ler Neyman (1937) para compreender como esta e outras técnicas se relacionam com a teoria da probabilidade.

Desvios moderados das premissas não resultam em grandes diferenças entre as técnicas de mínimos quadrados e verossimilhança máxima. Esta é a razão pela qual poucos pesquisadores se preocupam em usá-las. Não sabemos o formato real da distribuição dos resíduos dos valores de biomassa usados para construir a figura 55. Entretanto, assumindo uma distribuição de poisson, as estimativas por verossimilhança máxima dos parâmetros resultaram na posição da linha cheia mostrada na figura 60, que se sobrepõe quase completamente à linha obtida pelo modelo não linear mostrado na figura 58.

Os métodos de verossimilhança máxima minimizam a função de verossimilhança. Entretanto, esta função só pode ser determinada conhecendo-se a distribuição dos resíduos (Hilborn e Mangel 1997). As estimativas por verossimilhança máxima podem ser calculadas usando-se programas não-lineares (iterativos), minimizando o negativo do logarítmo da função de verossimilhança. Entretanto, há diferentes abordagens para a estimação por verossimilhança máxima (Bard 1974), e os resultados apresentam diferentes propriedades. Friendly (1995) apresenta algumas analogias físicas com a estimativa por verossimilhança máxima.

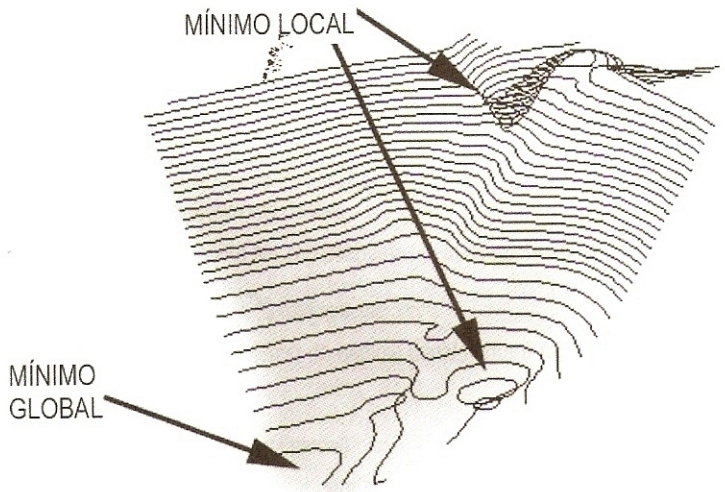
Uma vez que o modelo começa a ficar moderadamente complexo, não há um estimador de verossimilhança máxima simples e procedimentos muito complexos precisam ser programados no computador. É virtualmente impossível determinar a forma da distribuição estudada pela inspeção dos dados. Portanto, os métodos de verossimilhança máxima são os mais apropriados quando há fortes razões teóricas para se adotar um modelo em particular (Guttman 1985), e quando se tem grandes amostras. Não opte por usar estas técnicas, a não ser em colaboração com um estatístico muito competente. Quando se deparar com um modelo de verossimilhança máxima em suas leituras, deve interpretá-lo como faria com um modelo de mínimos quadrados, torcendo para que o autor tenha programado a função correta de verossimilhança. Muitos adeptos relatam que usaram verossimilhança máxima sem fornecer detalhes do método e suas premissas, de modo que freqüentemente é difícil, mesmo para estatísticos experientes, avaliar os resultados.

FIGURA 60

Problemas e armadilhas da estimativa não-linear

Todas as técnicas não-lineares (iterativas) apresentam uma fraqueza quando usada em modelos complexos. Os elementos de modelos simples podem ser descritos por uma reta ou por uma curva crescendo ou diminuindo monotonicamente. Entretanto, modelos complexos representam superfícies complexas, com ondulações. As técnicas iterativas geralmente aproximam a solução ótima, representada pelo ponto mais "baixo" na superfície, executando uma série de "pulos". Os pulos usualmente ficam menores a cada tentativa (iteração) para assegurar que o programa chegue a uma solução. Entretanto, este procedimento pode levar o programa a cair em algum local particular da superfície que apresente uma irregularidade chamada "mínimo local", ao invés de cair na porção realmente mais baixa de toda a superfície, o "mínimo global" (figura 61).

Para evitar isto, o pesquisador pode iniciar o processo repetidas vezes e escolher os valores iniciais para as estimativas, fazendo com que elas estejam próximas dos valores mais prováveis (i.e. fazendo com que o programa vá para a área onde provavelmente se encontra o mínimo global). Isto introduz um elemento de subjetividade no processo que alguns pesquisadores desaprovam. Entretanto, toda modelagem, inclusive a modelagem de rejeição de hipóteses nulas da estatística convencional, envolve um alto grau de subjetividade. É difícil lidar com modelos complexos, independentemente de com que método se trabalhe.

Os conceitos apresentados neste capítulo foram complexos e não temos espaço para explorá-los em detalhes. Entretanto, eles são conceitos puramente estatísticos, e não são tão importantes quanto os conceitos gerais que exploramos nos capítulos anteriores. No capítulo 10, lidamos com modelos lineares para descobrir como podemos construir modelos mais realistas com dependências entre variáveis independentes. Muitos programas de computadores usam métodos de verossimilhança máxima para estimar os parâmetros de equações lineares nestes modelos. Modelos complexos podem também ser atacados por técnicas não-lineares, mas usualmente, esta alternativa não é mais satisfatória do que os métodos convencionais.

FIGURA 61

CAPÍTULO 12

[A estatística multivariada é um vasto campo onde até os estatísticos experientes se movem cuidadosamente]

Análise multivariada: cortando as árvores para enxergar melhor a floresta

Não sem motivo, a estatística multivariada tem sido comparada à caixa de Pandora (James e McCulloch 1990). Na mitologia grega, Pandora era uma estátua dotada de dons divinos, mas que carregava uma caixa onde estavam guardados todos os malefícios da humanidade. A estatística multivariada é um vasto campo onde até os estatísticos experientes se movem cuidadosamente, mas apesar desta complexidade tem sido cada vez mais freqüente nos estudos ecológicos e é provavel que os biólogos jamais deixem de usá-la. Muitas das técnicas multivariadas e univariadas, incluindo as que discutimos até agora, podem ser vistas como casos especiais da análise de correlação canônica (Harris 1975). Entretanto, vamos nos concentrar em métodos que demandam menos matemática e tentar relacionar os padrões observáveis nos gráficos com aquilo que as técnicas estão tentando revelar. Nas seções anteriores, lidamos com relações que podiam ser tratadas matemática ou conceitualmente segundo a equação:

$$Y = a + b^{*}X_{\text{(possivelmente parcial)}}$$

A maioria das técnicas multivariadas usadas em ecologia podem ser representadas pela seguinte equação conceitual:

$$Y_1, Y_2, Y_3, \ldots Y_i = a + b^{*}X_{\text{(possivelmente parcial)}}$$

Entretanto, algumas vezes não há variáveis independentes. O objetivo da análise usualmente é reduzir o problema a alguma coisa como a seguinte equação:

$$Y' = a + b^{*}X_{\text{(possivelmente parcial)}}$$

onde Y' é uma ou mais variáveis conceituais compostas, destinadas a capturar a "essência" das múltiplas variáveis dependentes. O resultado pode ser representado em gráficos uni, bi ou tridimensionais.

Vamos começar com técnicas de ordenação, porque elas lidam com variáveis contínuas e são úteis para introduzir os conceitos que estão por trás da maioria das outras técnicas. A forma mais simples é a ordenação direta das variáveis dependentes por uma única variável independente. A tabela 13 mostra os dados para um estudo hipotético da distribuição de seis espécies de plantas (sp_i) em relação à precipitação média mensal de cada local. Os valores para cada espécie representam sua densidade (número por 100 m^2) em cada local. As espécies A e E são gramíneas (herbáceas), C e D são arbustos e B e F são árvores.

TABELA 13

LOCAL	spA	spB	spC	spD	spE	spF	Precipitação (mm)
1	9	1	0	0	0	0	270
2	0	1	0	0	0	1	315
3	0	0	0	0	4	0	200
4	0	2	1	1	0	2	255
5	4	0	0	0	6	0	190
6	0	1	0	0	0	0	290
7	8	0	0	0	7	0	150
8	6	0	0	0	3	0	125
9	2	0	2	2	3	1	230
10	0	2	1	1	0	1	290
11	0	0	4	4	1	0	240
12	10	0	0	0	1	0	100

Gráficos de gradientes

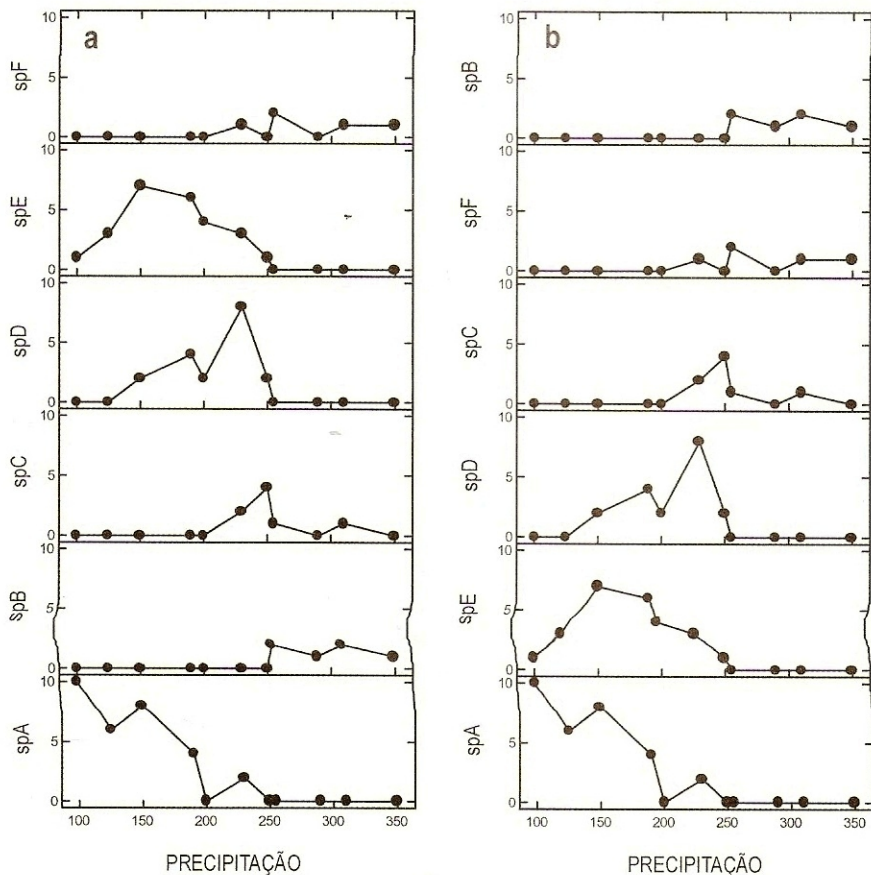
A figura 62a é um gráfico composto, mostrando a distribuição de cada espécie contra a precipitação pluviométrica. Notem que as diferentes espécies não têm uma distribuição similar ao longo do gradiente de precipitação. Algumas têm densidades maiores em locais com pouca chuva, outras apresentam densidades maiores com chuva intermediária e algumas têm densidades máximas (embora não muito grandes) em locais com muita chuva. A chuva parece estar influenciando as espécies, mas as relações de cada espécie com a chuva não são lineares. Entretanto, o padrão seria mais claro se tivéssemos apresentado os gráficos em uma ordem diferente, colocando os gráficos das espécies que tiveram as maiores densidades ocorrendo em áreas com grande precipitação acima e os gráficos das espécies com maiores densidades ocorrendo em locais menos chuvosos abaixo (figura 62b).

Agora ficou claro que a precipitação ordena as espécies. As maiores densidades formam uma diagonal ao longo do gráfico composto. Isto é chamado de "análise direta de gradiente", porque arranjamos as espécies em relação a um gradiente que esperávamos que fosse importante. Baseados na figura 62b, poderíamos dizer que cada espécie e a comunidade como um todo são organizadas ao longo da dimensão "precipitação".

Se não tivéssemos sido capazes de produzir um gráfico com as maiores densidades formando uma diagonal, concluiríamos que a chuva não era um fator determinante da estrutura desta comunidade. Diferentes tipos de estrutura irão refletir-se em diferentes padrões nos gráficos. Portanto, não devemos acreditar em uma estrutura (padrão), a menos que possamos visualizá-la em gráficos simples como a figura 62b.

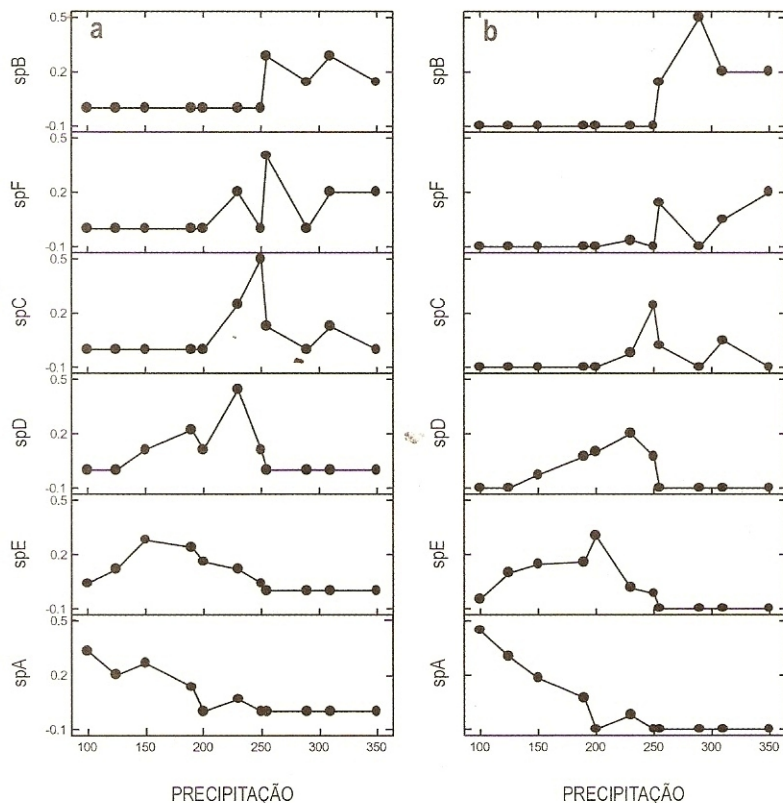
Parece estranho haver menos plantas nas áreas de maior precipitação, até levarmos em conta as diferentes formas de vida. Simplesmente não é possível empacotar tantas árvores na pequena área de um quadrado amostral como se pode fazer com gramíneas.

Medimos a composição da comunidade em termos do número de indivíduos, mas isto não reflete a biomassa, que provavelmente será muito maior nas árvores. Uma maneira de resolver este problema seria padronizar a densidade das

FIGURA 62

espécies. Se dividirmos o número de indivíduos de cada espécie encontrado em cada área amostral, pelo número total de indivíduos daquela espécie, todas as espécies estarão em uma escala adimensional equivalente, que varia de 0 a 1 (figura 63a).

O padrão agora ficou mais claro, com as árvores tendo tanto efeito quanto as gramíneas em nossa análise visual. Entretanto, nem sempre este será o caso. Algumas espécies podem ser raras simplesmente devido aos azares da amostragem, e atribuir um peso indevido para estas espécies, através desta padronização, pode obscurecer os padrões ao invés de revelá-los. Outra alternativa seria transformar os locais para ter as mesmas densidades de indivíduos, independente de espécies. Neste último caso, a análise visual não se preocupa com a densidade absoluta, mas enfatiza a densidade relativa (figura 63b). No entanto, sejam cuidadosos ao usar proporções ou outras transformações que limitem a amplitude dos dados, já que estes procedimentos podem criar padrões espúrios em algumas análises (Jackson 1997).

As transformações dos dados alteram as interpretações biológicas (Noy-Meir *et al.* 1975, Pielou, 1984, Johnson e Field 1993). Em nosso exemplo, o padrão geral permaneceu o mesmo, mas os detalhes de cada espécie mudaram. Na verdade, os gráficos 62b, 63a e 63b respondem questões diferentes. É importante compreender isto, antes de passarmos para a próxima seção. Muitas das técnicas de análise multivariada mais empregadas realizam transformações de dados internamente e o pesquisador deve estar certo de que estas transformações são apropriadas para a questão em pauta.

FIGURA 63

Gradientes hipotéticos

Até agora nossas análises se referiram a um gradiente ecológico que sabemos existir. Nossa questão era se existe um padrão associado com o gradiente que conhecemos. Entretanto, pode haver outros padrões, talvez até mais fortes, que estamos deixando de lado porque nos concentramos no gradiente de precipitação pluviométrica. Uma outra maneira de abordar a questão é perguntar se há padrões nos dados, independente de quaisquer padrões conhecidos. Isto é chamado de "análise indireta de gradiente". Há muitos tipos de análise indireta de gradiente, mas começaremos com uma que não requer muita matemática para ser compreendida. A esta altura do livro, esperamos que não se assustem mais com os nomes esdrúxulos pelos quais os estatísticos chamam suas análises. A que vamos tratar agora é chamada de "análise não-métrica de escalas multidimensionais" e é representada pela sigla NMDS. Considerando nosso exemplo, o objetivo da análise é descrever o padrão apresentado pelas seis espécies, em menos dimensões do que as apresentadas na tabela 13. Observe que aqui cada espécie representa uma dimensão, além, é claro, da dimensão representada pela precipitação. Uma das vantagens de NMDS é que podemos escolher para quantas dimensões queremos reduzir o nosso problema. Como sabemos que existe um gradiente de chuva e analisamos os dados em relação a esta dimensão nos gráficos anteriores, vamos fazer uma ordenação unidimensional dos "objetos", que em nossa análise são os locais, em relação aos "atributos", que em nossa análise são as espécies.

Ordenar os locais requer um passo intermediário. Não temos informação sobre os locais em relação a um gradiente externo. Temos apenas dados sobre o relacionamento entre os locais, baseados na distribuição de espécies. Então, perguntaremos o quanto os locais diferem considerando as espécies. Entretanto, esta questão nos leva a outra pergunta: qual será nossa medida de diferença? Diferenças em qual espécie? A medida mais simples de diferença (ou "distância") será apenas somar as diferenças entre os locais para cada espécie. Por exemplo, no local #1 contamos nove indivíduos da spA enquanto no local #2 não houve indivíduos desta espécie. Portanto estes locais diferiram por nove nesta variável. Ambos tiveram um indivíduo da spB, portanto a diferença foi zero, para esta espécie. A diferença média por espécies entre o local #1 e o local #2, somando todas as espécies foi 10/6 = 1,7. Se conduzirmos o mesmo procedimento para cada par de locais, obteremos o resultado mostrado na tabela 14, que é chamada de "matriz de associação".

Matrizes de associação podem ter medidas de similaridade ou de dissimilaridade e há muitas opções diferentes para medidas. Por exemplo, poderíamos ter calculado a correlação entre as contagens de espécies para cada par de locais. Isto teria resultado em uma matriz de similaridade dos locais. Mas, notem que o coeficiente de correlação teria implicado em uma transformação resultante da padronização por locais. Isto equivaleria a transformar os valores dentro de cada local para terem uma média igual a zero e o desvio padrão igual a um. A "distância" que usamos, chamada de distância "Manhattan" ou "city-block" não é a medida de associação mais usada e não é a que recomendaríamos para este tipo de dados. Entretanto, muitas das melhores medidas de associação são variações dela e ela é útil para ilustrar distâncias porque é intuitiva e fácil de calcular.

Os valores na diagonal da tabela 14 são todos zeros, porque a distância entre um local e ele mesmo é zero. Não preenchemos a porção superior direita da tabela porque ela seria o espelho da imagem inferior. Na tabela temos todas as informações a respeito de quão distante cada local está dos outros em relação à composição de espécies. Entretanto, isto ainda

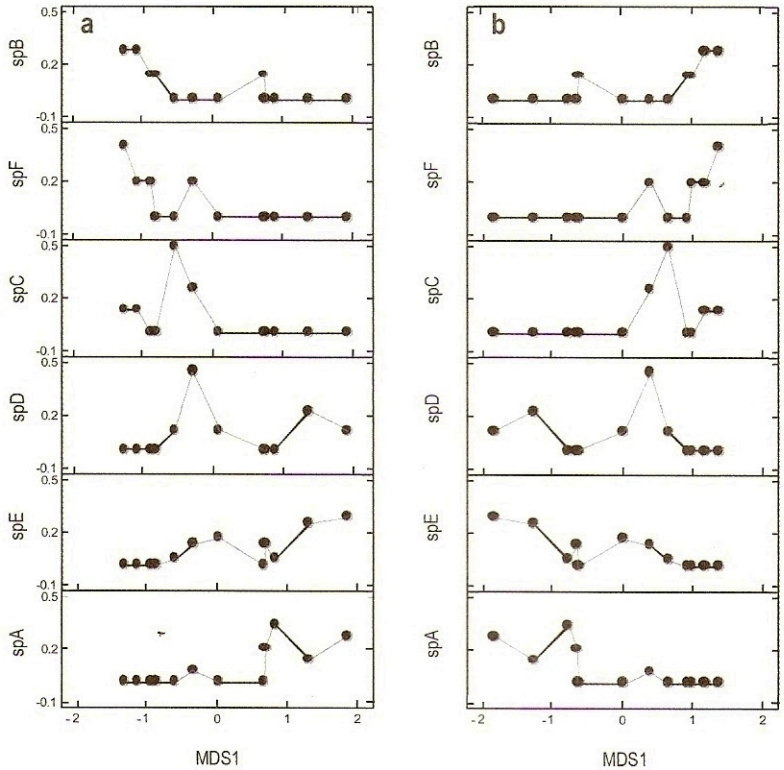
não simplificou nosso problema. Na verdade, esta tabela contém mais células que a original. Vamos usar a análise NMDS para ordenar os locais ao longo de um único eixo, de forma tal que, tanto quanto possível, as distâncias entre os locais ao longo do eixo sejam proporcionais às distâncias na tabela 14. O computador vai usar uma matemática complicada para reordenar (em nosso exemplo, em uma dimensão) os locais e verificar o quanto as distâncias entre eles ficaram proporcionais às da tabela 14. Em seguida, ele vai rearranjar os pontos e verificar se o resultado melhorou, até que não possa chegar mais perto das proporções originais. O leitor poderia fazer isto a mão, mas certamente será mais demorado do que o computador. Borg e Groenen (1997) fornecem um texto razoavelmente digerível sobre como o computador pode ser programado para fazer isto. Entretanto, iremos apenas comparar a ordem do computador com nossa análise direta de gradiente, já que estamos na situação pouco usual de conhecer com antecedência o gradiente ecológico "real". A figura 64a apresenta os locais ao longo do novo eixo criado pela análise NMDS, e chamamos este eixo de MDS1.

TABELA 14

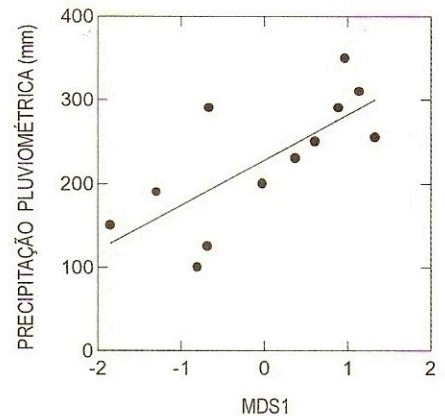
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
#1	0											
#2	1,7	0										
#3	2,3	1,0	0									
#4	2,3	0,7	1,7	0								
#5	2,0	2,0	1,0	2,7	0							
#6	1,5	0,2	0,8	0,8	1,8	0						
#7	1,5	2,8	1,8	3,5	0,8	2,7	0					
#8	1,2	1,8	1,2	2,5	0,8	1,7	1,0	0				
#9	2,7	1,7	1,3	1,7	1,7	1,8	2,5	1,5	0			
#10	2,2	0,5	1,5	0,2	2,5	0,7	3,3	2,3	1,5	0		
#11	3,2	1,8	1,8	1,8	2,8	1,7	3,7	2,7	1,5	1,7	0	
#12	0,5	2,2	2,2	2,8	1,8	2,0	1,3	1,0	2,5	2,7	3,0	0

Notamos que na figura 64a há um padrão diagonal que corre na direção oposta à da figura 63. A direção do novo eixo é completamente arbitrária, porque o eixo foi desenhado somente para manter as distâncias relativas entre os pontos. Portanto, podemos reverter o eixo MDS1 (figura 64b).

O padrão é similar ao obtido com a análise direta de precipitação, mas a análise MDS não foi capaz de reproduzir o padrão original completamente. Ainda assim, MDS1 representou a precipitação razoavelmente bem, embora nenhuma observação direta de chuva tenha sido usada na ordenação. A análise fez isso usando somente as similaridades (mais corretamente, as dissimilaridades) entre os locais em termos de contagens de plantas. Se colocarmos em um gráfico o gradiente verdadeiro (precipitação) contra o gradiente predito pela análise NMDS, veremos que o eixo hipotético conseguiu predizer cerca de 50% ($r^2 = 0.53$) da variação da precipitação pluviométrica (figura 65).

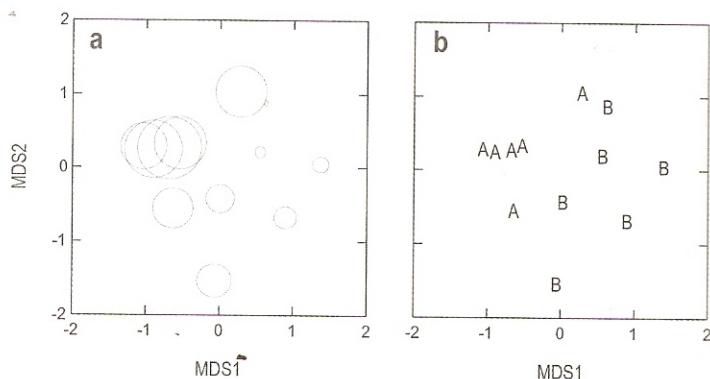
FIGURA 64

Isto foi bem razoável, considerando que nenhuma das distribuições de espécies ao longo do gradiente de chuva (figura 62b) formava uma curva muito suave e que a associação das variáveis originais com a precipitação não foi mais forte do que a esperada para dados ecológicos.

Ordenações indiretas sempre seguem os mesmos passos, embora isto nem sempre seja óbvio a partir dos programas. Primeiro os dados podem ser transformados e o tipo de transformação deveria ser dependente da questão formulada. A seguir, uma matriz de associação das distâncias (ou das similaridades) entre os objetos é construída. A matriz de associação é baseada em uma medida obtida a partir da média de todos os atributos. Portanto, há apenas uma distância entre cada par de objetos analisados, independente de quantas variáveis (atributos) sejam medidas. O programa então arranja os objetos ao longo de um ou mais eixos (usualmente dois eixos ou dimensões) que melhor refletem os padrões encontrados nos dados. Com sorte, estes eixos refletirão as variáveis ecológicas que causaram os padrões nos dados. Na melhor das hipóteses, esperamos somente uma aproximação grosseira das dimensões reais que determinaram os padrões e se escolhemos uma

FIGURA 65

transformação imprópria, ou uma medida de associação ou técnica de ordenação inadequada, pode não haver correspondência entre os eixos derivados e os gradientes ecológicos reais (procurem Kenckel e Orloci 1986 para exemplos). Algumas vezes, o pesquisador está interessado em uma ordenação dos atributos ao invés dos objetos e alguns métodos podem fazer ordenações simultâneas de objetos e atributos. Entretanto, a lógica da análise permanece a mesma.

Mais do que uma dimensão

Poderíamos ter pedido ao programa de NMDS para arranjar os locais em duas dimensões. Sabemos que, na verdade, há apenas um gradiente nestes dados, porque nós os criamos. Entretanto, ainda assim podemos ver o padrão em duas dimensões. Fazendo o símbolo dos locais ser proporcional à chuva (figura 66a) ou nomeando os locais com chuva abaixo da média como "B" (baixa precipitação) e acima da média como "A" (alta precipitação), podemos ver a associação entre precipitação e locais em termos de composição de espécies.

FIGURA 66

Se tivéssemos incluído muitos eixos, degradaríamos o padrão. Há uma grande discussão na literatura sobre a seleção do número de eixos que devem ser incluídos, especialmente para análises de vetores "*eigen*" (p. ex. Jackson 1993), das quais trataremos logo adiante. Entretanto, é muito raro que um padrão possa ser discernível em mais do que duas dimensões, e permanecer com muitas dimensões contraria o propósito primário da análise, que é reduzir a dimensionalidade (Gauch 1982a, James e McCulloch 1990). Freqüentemente, os eixos não são "significantes" nos critérios internos da análise, mas ainda assim podem ser carregados de significado em termos dos gradientes externos (Gauch 1982b, Økland, 1999).

[Há muitas maneiras de executarmos a análise de escala multidimensional (MDS), que podem ser métricas, não-métricas ou um híbrido dos dois (p. ex. Borg and Groenen 1997, Faith, Minchin and Belbin 1987).]

Freqüentemente se diz que os eixos resultantes de MDS não têm interpretação lógica. É verdade que podemos girar os eixos para qualquer posição no plano dos dados, sem mudar as distâncias relativas entre os pontos. O programa que usamos girou os eixos para obter a maior correlação entre as

variáveis originais e o primeiro eixo MDS. Entretanto, isto não faz sentido necessariamente, já que as distâncias entre os pontos não são derivadas de correlações. O gradiente de precipitação parece correr como uma diagonal através do espaço bidimensional determinado pelos eixos MDS. Há um outra classe de técnica de ordenação, baseada na análise de "*eigen*", cujos eixos resultantes podem ser matematicamente relacionados com os valores originais.

Análises de vetores *"eigen"*

Estabelecendo-se algumas premissas importantes, e usualmente improváveis, é possível determinar a posição dos eixos no espaço multidimensional usando-se a álgebra de matrizes, ao invés do método de "tentativa e erro" das técnicas de iteração. As análises baseadas neste princípio são chamadas de "análises de vetores *eigen*" pelos estatísticos e algumas vezes têm sido traduzidas para o português como "análises de auto-vetores". "*Eigen*" é uma palavra em alemão que significa "característica" ou "próprio". Os eixos das "análises *eigen*" são derivados diretamente dos atributos (características) dos objetos.

A primeira técnica de ordenação baseada em vetores "*eigen*" a ser amplamente usada foi a análise de componentes principais (PCA), que geralmente usa o coeficiente de correlação de Pearson como medida de associação (Pearson 1901). As outras técnicas são variações desta. Por exemplo, a análise de correspondência (CA) pode ser calculada da mesma maneira que a PCA, depois de se substituir a distância do coeficiente de correlação pelo Qui-quadrado e padronizar os objetos pela amplitude. Se alguma outra distância é usada, e a matriz de associação for transformada para possuir propriedades métricas, a análise é chamada de análise das coordenadas principais (PCoA). PCA e CA são populares, porque são fáceis de serem calculadas, os eixos resultantes podem ser matematicamente relacionados com as medidas originais e é possível projetar tanto os objetos quanto os atributos no mesmo espaço de coordenadas.

Algumas das vantagens das "análises *eigen*" são ilusórias. PCA assume relações lineares entre as variáveis, o que pode não existir. Se os objetos são padronizados pela divisão pela soma antes de se executar a análise, os eixos da PCA serão artefatos (Jackson 1977).

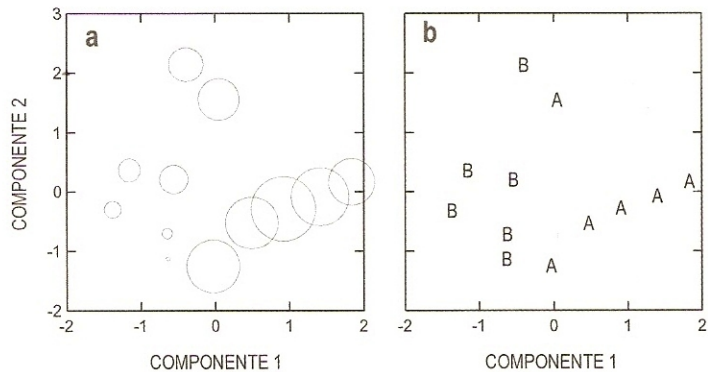
Para gradientes longos, nos quais a maioria dos atributos vão possuir vários valores iguais a zero, as análises *eigen* tendem a distorcer as relações entre os objetos e vão indicar formas de ferradura para gradientes lineares. Uma variação da CA, que poderia ser traduzida como "análise de correspondência destendenciada" *(Detrended Correspondence Analysis - DCA)* foi proposta para corrigir estes efeitos de "ferradura", mas provavelmente transforma em retas também as relações que são verdadeiramente curvilíneas no espaço ecológico (p. ex. Wartenberg *et al.* 1987). Kenckel e Orlocci (1986) fornecem exemplos do que DCA pode fazer com gradientes conhecidos. Além disso, DECORANA, o programa de computador que tornou CA e DCA tão popular, e que foi usado pela maioria dos autores que publicaram usando estas técnicas, tinha um erro (Oksanen e Minchin 1997). Isto é preocupante e ilustra porque é importante analisar criticamente os resultados de análises multivariadas em relação aos dados originais, antes de depositar muita fé neles. Para análises em um único plano (duas

dimensões), De'Ath (1999) mostrou que usar análises de curvas principais (*Principal Curves Analysis*) para suavizar os resultados obtidos por outras técnicas de ordenação, pode recuperar melhor os gradientes ecológicos. Entretanto, isto vai muito além do escopo deste livro introdutório.

A relação aparente entre as variáveis originais e os eixos derivados raramente é simples, para quaisquer das análises. As aparentes relações geométricas só são reais se uma medida geométrica foi usada, como a distância euclidiana, se as variáveis não foram transformadas, e se todas as variáveis são realmente mensuráveis na mesma escala. Se variáveis ambientais como pH, temperatura, condutividade, velocidade da água e granulometria do sedimento foram usadas, é ridículo inferir que todos estes atributos podem ser combinados geometricamente em qualquer sentido. Mesmo densidades de plantas não são comparáveis, a não ser entre espécies de forma de vida e tamanhos similares. Combinar números de gramíneas, arbustos e árvores pode ser tão esotérico quanto combinar temperatura e velocidade da água.

A despeito das diferenças metodológicas e bases conceituais, o padrão capturado pela ordenação MDS (figura 66) e os dois primeiros eixos da PCA (figura 67) foram similares.

FIGURA 67

Se o padrão nos dados for muito forte e causado por poucos gradientes, todos os métodos comumente usados fornecerão resultados semelhantes. Entretanto, as diferenças podem ser importantes. Não recomendaríamos as análises apresentadas acima para os dados da tabela 13. Sabemos que outras medidas de associação e/ou transformações serão mais apropriados para a maioria dos estudos ecológicos. Nós apresentamos estes métodos apenas por motivos didáticos, como uma introdução simplificada das técnicas. A única maneira de se escolher o melhor método para uma determinada situação é considerar o provável padrão que o gradiente ecológico irá impingir aos dados (p.ex. Kenckel e Orloci 1986, McCune e Grace 2002). Se o leitor ainda não for capaz de desenhar um gráfico hipotético ou tabela que ilustre os padrões que podem estar em seus dados, deveria se empenhar em desenvolver esta habilidade antes de proceder o próximo passo e procurar por padrões. Gauch publicou um livro de grande lucidez sobre técnicas multivariadas, em 1982. A maioria de sua discussão sobre a técnica de ordenação apropriada está ultrapassada e suas recomendações diferem de muitas que fizemos aqui. Entretanto, sua discussão dos usos das técnicas multivariadas e dos padrões em tabelas de dados são tão atuais quanto eram vinte anos atrás. Seu capítulo introdutório ainda é um dos melhores apanhados das técnicas multivariadas na ecologia e deve nortear qualquer curso de técnicas multivariadas para ecólogos.

Nenhuma técnica é adequada para todos os propósitos. Por exemplo, Azevedo-Ramos *et al.* (1999) usaram MDS e a associação Bray-Curtis para ordenar uma comunidade de girinos. No mesmo trabalho, usaram MDS com a medida de associação de Gower para ordenar a comunidade de predadores de girinos e PCA para ordenar as variáveis ambientais. Cada uma foi escolhida para refletir um tipo diferente de padrão. Entretanto, não podemos entrar em detalhes agora sem perder de vista os conceitos gerais que vamos precisar para considerar outros métodos multivariados.

A força da cultura: testes de significância

Agora que já têm alguma idéia de o que as "dimensões" são em relação aos seus dados (talvez "eixos fantasmas" seja um termo mais apropriado), podemos retornar a algumas técnicas de ordenação diretas, que usualmente são usadas apenas para avaliar a "significância" de gradientes conhecidos. A mais geral é uma análise multivariada, análoga a regressão múltipla, chamada regressão múltipla multivariada (RMM). Considere a seguinte equação:

$$Y_1, Y_2, Y_3, \ldots Y_i = a + b_1{}^*X_1 + b_2{}^*X_2 + \ldots\ldots b_i{}^*X_i$$

A RMM pergunta se há uma combinação linear de variáveis do lado esquerdo da igualdade que pode ser explicada por uma combinação *linear* das variáveis do lado direito. Notem nossa ênfase em *linear*. Se a estrutura dos dados do lado esquerdo não pode ser capturada por PCA, não deveríamos usar a RMM. Há muitos testes estatísticos multivariados e cada qual tem diferentes premissas. Em todo o caso, a interpretação de um resultado "significativo" não é tão fácil, porque ele pode ter aparecido devido a uma única variável ou a uma combinação de variáveis. Mesmo que encontremos algo "significativo", ainda será necessário muita exploração de dados para descobrir o que aconteceu. Provavelmente, é por isso que é tão raro encontrar RMM sendo usada na literatura ecológica.

Quanto aos nossos dados, onde temos apenas uma variável explanatória, as três estatísticas multivariadas mais comuns (lâmbda de *Wilksons*, *Pillai Trace* e *Hotelling-Lawley Trace*) rejeitam a hipótese nula de ausência de efeito de precipitação a $P = 0{,}070$. Havíamos perguntado o quanto a precipitação explicou a variação das variáveis dependentes. Entretanto, uma questão diferente, mas igualmente válida, seria "a precipitação explica o padrão mais forte nas variáveis dependentes" ? Sabemos que há apenas um gradiente direcionando estes dados, mas os leitores não saberão disto com seus dados de campo. Assumindo que os dois eixos MDS são a melhor representação do padrão mais forte nas variáveis dependentes, podemos fazer uma análise de RMM sobre as duas variáveis MDS. Isto rejeita a hipótese nula de ausência de efeito de chuva a $P = 0{,}023$.

Ambas as questões foram válidas, mas a diferença é importante. A segunda questão, que perguntou apenas sobre o efeito de chuva sobre o padrão mais marcante foi capaz de rejeitar a hipótese nula com mais confiança. Geralmente isto acontece. Por se usar uma ordenação indireta para reduzir a dimensionalidade, pudemos nos focar no padrão ao invés de na variabilidade total e fomos mais capazes de identificar um efeito. Isto não aconteceu por o MDS ser mais hábil em manipular

relações não-lineares do que as análises "*eigen*". Se aplicarmos a RMM sobre os dois primeiros eixos da PCA, rejeitaríamos a hipótese nula a $P = 0{,}027$. As análises dos padrões mais fortes simplesmente são mais capazes de detectar efeitos que as análises que usam toda a variabilidade. O princípio é similar ao uso de PCA para produzir variáveis lineares aditivas para análise de regressão (p. ex. Short *et al.* 1983). Gauch (1982a,b) chamou este processo de "redução de ruído por ordenação de vetores *eigen*", mas nossa ordenação MDS (que não usa vetores "*eigen*") também foi efetiva. Lembrem-se também que muito da variabilidade não explicada pode ser devido à tentativa de ajustar um modelo impróprio, ao invés de apenas resultado de "ruído" aleatório (Økland 1999).

A esta altura, esperamos que os leitores estejam se perguntando porque categorizamos uma variável contínua nas figuras 66b e 67b. Com certeza não recomendamos que se faça isto com freqüência, embora em alguns casos possa ser útil para revelar rapidamente padrões em gráficos tridimensionais, como os das figuras 66 e 67, especialmente quando seus leitores ou ouvintes não dispuserem de muito tempo, como no caso de apresentações em seminários. Apenas para própositos de ilustração, vamos assumir que os níveis de precipitação baixa e alta sejam realmente categóricos e aplicar uma análise multivariada análoga a ANOVA, usando estes níveis. Todas as precauções exigidas para a ANOVA se aplicam para a "análise de variância multivariada" (MANOVA). Na verdade, ela é apenas um caso especial da RMM, que usa variáveis substitutas (*dummy variables*) para codificar as categorias.

Há muitas estatísticas que podem ser usadas para avaliar a significância de MANOVA, mas simulações indicam que uma delas denominada *"Pillai trace*" é a mais robusta à violações das premissas (Olson 1976, Johnson e Field 1993). Um teste não-paramétrico análogo a MANOVA, que usa aleatorização para gerar valores de probabilidade (Anderson 2001) pode ser o mais apropriado em muitas situações. Aplicando MANOVA em nosso exemplo, rejeitamos a hipótese nula de ausência de efeito dos níveis alto e baixo sobre combinações lineares das variáveis dependentes a $P = 0{,}021$. De novo, se perguntássemos apenas a respeito dos maiores padrões e usássemos a MANOVA para testar o efeito dos dois níveis de precipitação sobre os eixos MDS, teríamos um teste mais poderoso e rejeitaríamos a hipótese nula a $P = 0{,}002$.

Há vários testes para examinar os efeitos de variáveis contínuas e categóricas que operam diretamente sobre a matriz de associação (lembrem-se que nossa matriz de associação encontra-se na tabela 14). Estes testes são derivados do teste de Mantel (Mantel e Varland 1970), que é um teste de permutação similar em seus princípios ao nosso teste de *DIF* (capítulo 5). Eles usam mais informações do que há nos eixos de ordenação, e quando há apenas um gradiente sendo testado, podem ser mais poderosos para detectar os gradientes conhecidos. Em nosso exemplo, testes de permutação para um efeito de precipitação como uma variável contínua ou como variável categórica com dois níveis, rejeitam a hipótese nula a $P < 0{,}001$ e $P = 0{,}006$, respectivamente. Entretanto, quando a variável categórica apresenta mais de dois níveis (p. ex. Luo e Fox 1996), ou há mais que uma variável independente (Anderson e Legendre 1999) eles, algumas vezes, têm menos poder do que as análises dos padrões derivados por ordenação indireta. Observem que a técnica recomendada por Smouse *et al.* (1986) provavelmente tem um erro do tipo I inaceitavelmente alto (Anderson e Legendre 1999).

Peres-Neto e Jackson (2001) descreveram um teste baseado na "sobreposição de Procrustes" (alinhamento simultâneo em várias dimensões) que potencialmente usa todas as distâncias originais, e que é mais poderoso do que o teste de Mantel. No entanto, Peres-Neto e Jakson (2001) usaram esta técnica depois de reduzir a dimensionalidade por ordenação. Neste caso, somente os principais padrões seriam investigados e a técnica produziria resultados semelhantes aos de algumas outras descritas nas seções anteriores.

Como em outras análises de gradientes diretos, os testes de permutação de matrizes de associação perguntam se há alguma associação, e não se os fatores contribuem para os maiores padrões. Eles também têm a deficiência de só testar por padrões nos dados que estão associados com a(s) variável(eis) independente(s) e não descrevem o padrão. Entretanto, extensões do conceito para detectar padrões espaciais podem ser muito úteis (p. ex. Fortin e Gurevitch 1993, Legendre 1993, Sawada 1999). Como em outras técnicas estatísticas, estes testes têm uso limitado para refinar nossos modelos biológicos, se os usarmos apenas para gerar valores de probabilidade. Eles testam as hipóteses, mas não fornecem uma indicação real do porquê elas são rejeitadas ou não.

Um método aparentemente atraente de análise direta de gradientes é a Análise de Correspondência Canônica (CCA), que coloca os objetos, atributos e as variáveis preditivas todos em um mesmo gráfico (ter Braak 1986). Ela parece fazer tudo que desejamos. Entretanto, há duas maneiras diferentes de se usar esta análise (MacCune 1997). Se o leitor usá-la como uma análise direta de gradiente, ela ignorará os principais padrões que não estiverem linearmente associados com os gradientes esperados, que poderiam ser encontrados por análises indiretas de gradiente (Økland, 1996). Se a usarmos em sua forma original, na verdade será apenas uma análise indireta de gradiente seguida de um teste de associação das variáveis independentes ("ambientais") com os gradientes derivados (MacCune 1997). Poucas pessoas realmente entendem a diferença entre uma análise direta de gradientes e uma análise indireta de gradientes. Legendre e Legendre (1998) explicaram os passos necessários para se obter a ordenação mas a álgebra envolvida está além da capacidade de programação da maioria dos biólogos. A lógica do CCA é mais facilmente compreendida depois de se entender a lógica de uma técnica estreitamente relacionada ao CCA: a "Análise de Redundância" (RDA).

A primeira aplicação da técnica é atribuida a Rao (1964), mas a melhor descrição da técnica pode ser encontrada no livro de Legendre e Legendre (1998). Estes autores usaram regressão múltipla para modelar as relações lineares entre as variáveis independentes e as variáveis preditoras. Os dados originais são substituídos pelos valores previstos pela regressão múltipla e as ordenações são feitas usando estes valores derivados. Isto garante uma relação linear entre os atributos, as variáveis preditoras, e os eixos derivados, mas as ordenações resultantes podem ter pouca relação com os padrões existentes nos dados originais.

Ambos CCA e RDA têm a premissa de que as relações entre as variáveis dependentes e independentes são lineares. Para contornar este problema, Makarenkov e Legendre (2002) desenvolveram RDA e CCA não lineares baseadas em regressão polinomial, mas a interpretação dos vetores resultantes é ainda mais complicada. Seu programa é disponível gratuitamente

pela internet, na página do "Ecological Archives E083-018-S1.

[http://*www.esapubs.org/archive/*.]

Legendre e Anderson (1999) transformaram distâncias não-métricas para fazê-las métricas e empregaram RDA em um delineamento fatorial, para relacionar variáveis independentes ao eixos PCoA. Em seguida, eles avaliaram a significância das estatísticas com testes de permutação. Esta abordagem permite testes para interações em uma situação experimental. Embora a análise produza um eixo PCoA a menos do que o número de objetos, independentemente do número de atributos, ela não é usada primariamente para reduzir o número de dimensões. Ao contrário, ela pode ser considerada uma transformação para produzir variáveis que preenchem as premissas da análise de alocação de variâncias. Legendre e Anderson (1999) fornecem exemplos que ilustram como a técnica permite interpretações das interações bem como dos principais efeitos que são observados em uma análise indireta de gradiente (MDS).

Legendre e Gallagher (2001) mostraram como transformações dos dados originais permitem que muitas medidas de distâncias sejam representadas por distâncias euclidianas, aumentando consideravelmente as possibilidades para análises baseadas em PCA, PCoA ou RDA.

Méot *et al.* (1998) lançaram mão de matemática ainda mais complicada para repartir variação espacial e ambiental. Entretanto, é preciso extremo cuidado quando variáveis fantasmas vão substituir variáveis reais (Johnson e Field 1993). Pode ser muito difícil determinar o quê realmente mudou. Lembrem-se que outro conjunto de dados coletado do mesmo sistema poderá gerar eixos diferentes, não importa qual técnica seja usada. Se o leitor tiver a sorte de dispor de muitas observações independentes, pode decidir dividir seus dados em um subconjunto de exploração e outro de validação (p. ex. Hallgren *et al.* 1999).

Recentemente, alguns pesquisadores vêm propondo alocar variâncias à partir da matriz de distâncias, sem correções ou análises de eigen (McArdle and Anderson 2001), ou alocá-las diretamente do espaço multidimensional das variáveis originais (Anderson 2001), mas estas técnicas são muito complexas para serem tratadas aqui.

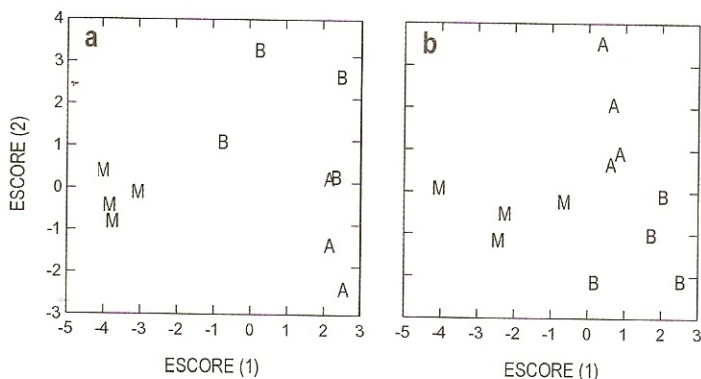
Ao fazer inferências sobre as variáveis produzidas a partir de técnicas multivariadas, o pesquisador que não for cauteloso estará apenas substituindo inferências fracas sobre variáveis fortes por inferências fortes sobre variáveis fracas. A interpretação destas análises é uma arte. Portanto, se o leitor não estiver se sentindo especialmente artístico após ler os parágrafos anteriores, é aconselhavel evitar estas técnicas.

Discriminando entre grupos

Uma outra técnica para testar diretamente gradientes é a "Análise de Função Discriminante" (DFA). Este método é estreitamente relacionado à MANOVA, e tenta determinar a combinação linear de variáveis que melhor discriminam entre dois ou mais grupos. Ela tem todas as premissas usuais de análises que repartem a variância (Williams 1983). Adicionalmente, a maioria dos estudos ecológicos publicados com esta análise tiveram tamanhos amostrais insuficientes (Williams e Titus 1988). DFA pode ser usada para produzir gráficos que se assemelham aos obtidos por ordenação indireta. Entretanto, estes gráficos tem uma interpretação muito diferente (Manly 1997). Vamos ilustrar isto com nossos dados da tabela 13, estabelecendo 3 grupos em relação à precipitação; baixa (B) quando for menor ou igual a 190 mm; média (M), quando estiver compreendida ente 191 e 260 mm; e alta (A), quando a precipitação for maior ou igual a 261 mm . Agora vamos aplicar a análise discriminante e ver se ela consegue discriminar os objetos (locais) em função de seus atributos (vegetação) (figura 68a).

A separação em baixa, média e alta precipitação parece razoável e o programa informa que classificou corretamente todos os locais e fornece as equações que alocam os objetos nas categorias, baseado em seus atributos. Entretanto, observe que o gráfico da figura 68a não é equivalente às ordenações indiretas nas quais a variável independente não foi usada para posicionar os locais. A figura 68b mostra a mesma análise com os dados de espécies substituídos por dados produzidos por um gerador de números aleatórios. A separação parece até melhor e o programa classificou corretamente todos os locais com relação às categorias de "precipitação". A interpretação dos padrões nos gráficos produzidos por CCA e RDA sofrem das mesmas limitações que a interpretação dos gráficos produzidos por DFA. Todas estas análises são baseadas em análises diretas de gradientes nas quais os padrões são forçados a coincidir ao máximo com as variáveis preditoras.

FIGURA 68

A análise sempre encontra combinações lineares das variáveis para separar as categorias, especialmente se o número de observações é pequeno, o que é típico de experimentos ecológicos (Williams e Titus 1988). Não deposite confiança em equações para separar categorias que foram geradas por DFA, a menos que elas tenham sido validadas por um conjunto independente de dados, ou pelo menos por um procedimento baseado em randomização como o *"Jack-Knife"* (Manel *et al.* 1999).

Categorias que se ramificam como árvores

Algumas vezes, a associação entre objetos pode ser expressa como distâncias ao longo de ramos de uma árvore. O exemplo mais comum é o de relacionamentos filogenéticos. A maioria das hipóteses modernas sobre filogenia são baseadas em análises cladistas e estas estão longe de serem simples. Os atributos são ordenados lógica ou estatisticamente, de forma que as distâncias entre as *taxa* são baseadas apenas no que se acredita serem características derivadas compartilhadas (sinapomorfismos) (Rohlf & Sokal 1981). Se acharam os conceitos de matrizes de associação e várias medidas potenciais de distâncias ecológicas complicado, então se preparem para o que vem com as análises cladísticas. Este é um campo enorme e complexo, e não cabe em nosso texto resumido. Entretanto, os leitores podem ser forçados a fazer uso da literatura desta área, especialmente se tiverem problemas com pseudo-repetições filogenéticas (capítulo 5). Aqui vamos considerar apenas diagramas (dendrogramas) que são baseados em distâncias ecológicas simples, como as usadas em análises de ordenação.

Árvores podem ser construídas em forma de chaves dicotômicas e existem alguns programas de computador para automatizar este processo (p. ex. Árvores de Regressão ou *Regression Trees*). Entretanto, a maioria das análises usa uma matriz de associação como a da tabela 14 para determinar as distâncias entre os objetos e entre as bifurcações (nós) das árvores. O processo pode ser aglomerativo, adicionando-se objetos um por vez para formar grupos, ou divisivo, separando as espécies hierarquicamente em subgrupos, até que cada subgrupo se constitua de apenas uma espécie ou objeto. Os membros do grupo podem ser definidos pela distância ao centro do grupo, a distância ao membro mais próximo, a distância ao último membro, e ainda de outras maneiras. Novamente, estamos procurando por padrões na matriz de dados, e não faz sentido procurar por estes padrões a não ser que possamos imaginar como eles deveriam aparecer em uma matriz de dados hipotéticos.

A escolha da técnica de agrupamento é provavelmente mais difícil do que a escolha entre as técnicas de ordenação, e só pode ser feita baseando-se na estrutura esperada dos dados. A opção de tentativa e erro é inaceitável. Se usarmos todas as combinações possíveis de medidas de distâncias e técnicas de agrupamento, é quase garantido que poderemos produzir um agrupamento que se conforme a qualquer hipótese ecológica que formulemos. Portanto, ficar procurando entre as diferentes técnicas até encontrar aquela que forneça os resultados que você esperava é usar a estatística de um modo inadequado.

Alguns pesquisadores recomendam que as técnicas de agrupamento sejam usadas junto com técnicas de ordenação para uma análise preliminar dos dados (p. ex.Gauch 1982a, Belbin 1992). Isto é apropriado se o leitor pretende permitir que o cómputador gere hipóteses por si, mas isto raramente é eficiente. As árvores geradas por análises de agrupamento *(cluster analysis)* parecem ser em duas dimensões, mas isto é ilusório. Eles representam móbiles, que podem ser girados em qualquer ângulo em cada nó. Eles têm um grande apelo sobre nós, porque parecem categorizar o mundo, e geralmente nos sentimos mais confortáveis com categorias do que com variáveis contínuas mais realistas (capítulo 7). As análises de ordenação podem revelar agrupamentos (*clusters*) porque estes são formados por gradientes abruptos. Entretanto, as análises de agrupamento são geralmente ineficientes para revelar gradientes suaves.

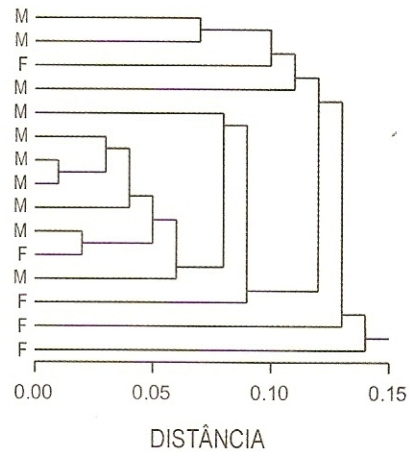
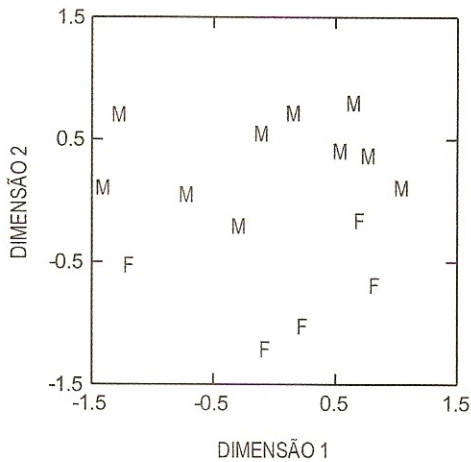
As análises de agrupamento só percebem agrupamentos de certos formatos no espaço multivariado. Vamos ilustrar isto com dados de medidas biométricas dos estudantes de nosso curso de estatística. As variáveis medidas foram os diâmetros da cabeça, pescoço, peito, cintura, quadril e comprimento da perna. As análises de agrupamento são análises indiretas de

gradientes, mas só podemos avaliá-la se usarmos dados com gradientes conhecidos. A experiência nos ensina que seres humanos nas categorias de sexo "masculino" (M) e "feminino" (F) diferem em forma, portanto vamos usar análises de ordenação e agrupamento para verificar o quanto podemos identificar esta diferença. A análise de agrupamento misturou as cinco mulheres com os homens (figura 69). As mulheres tenderam a ser consideradas muito diferentes umas das outras (distâncias longas ao longo dos "galhos" do dendograma) e ocorreram em todos os cinco principais agrupamentos diferentes (nós acima de 0,10). Baseado nesta análise, provavelmente descartaríamos a hipótese de que o sexo afeta a forma das pessoas.

FIGURA 69

Antes do agrupamento, fizemos uma transformação para padronizar os dados, dividindo o valor de cada medida pela soma das medidas de todos os atributos de cada estudante, porque estávamos interessados em descobrir o quanto a análise poderia detectar as diferenças na *forma* devido ao sexo. Como a soma de atributos varia entre indivíduos, se não fizéssemos esta padronização, as diferenças no *tamanho* das pessoas afetariam o padrão. Com a padronização por pessoa (os objetos), a soma para cada pessoa fica igual a um e o tamanho geral não pode afetar o padrão. Se não tivéssemos padronizado, a análise poderia ter capturado diferenças simplesmente devido ao tamanho dos objetos (estudantes). Nesta análise, como em todas as outras análises multivariadas, as transformações de dados mudam a questão que está sendo respondida. Usando a mesma transformação e a mesma medida de distância, uma análise MDS bidimensional produziu um padrão que claramente indica que homens e mulheres geralmente diferem em forma, muito embora alguns indivíduos estejam mais perto do sexo oposto do que do próprio sexo (figura 70).

FIGURA 70

A análise de gradiente (variáveis contínuas) pode detectar melhor o padrão do que a análise que tentou criar categorias, ainda que o padrão que estivéssemos procurando pudesse ser descrito por categorias. Todas as distâncias comumente

empregadas e as técnicas de agrupamento e de ordenação fornecem resultados qualitativamente similares para este exemplo. O fato de a variável categórica "sexo" não poder explicar em detalhes as diferenças entre nossos objetos (estudantes) poderia nos levar a procurar por variáveis contínuas, como nível de hormônios ou idade, que poderiam ser melhores preditores das diferenças entre os estudantes.

Se o leitor desejar determinar o número de agrupamentos apropriado para um conjunto de dados, precisará conhecer muito sobre a estrutura dos dados (Milligan e Cooper 1985, Dale 1988). Se decidir que pode prever a forma de seus agrupamentos *a priori*, e portanto pode selecionar um algorítmo de agrupamento apropriado, pode usar métodos de randomização, como a técnica de "*bootstrap*", para ajudar a identificar o número de agrupamentos naturais (Pillar 1999). Se quiser decidir simultaneamente o melhor número de agrupamentos e estimar os parâmetros que descrevem estes agrupamentos, considere o uso do "agrupamento de mensagem de menor comprimento" (*minimum message length clustering*), uma técnica emprestada da teoria de informação que parece ser promissora (Wallace e Dowe 2000). Pode também combinar testes estatísticos com análises multivariadas de *cluster* em uma análise chamada "árvores de regressão" (De'Ath 2002). A quantidade de análises potenciais é muito grande, e uma seleção somente poderá ser feita baseada num profundo conhecimento dos processos que determinam os padrões em um determinado conjunto de dados.

Selecionando variáveis

A seleção de variáveis é uma dificuldade tanto nas análises multifatoriais, como para as análises multivariadas. Os estudantes freqüentemente usam os mesmos critérios sugeridos para a seleção de variáveis em regressão múltipla. Normalmente eles excluem as variáveis que são altamente correlacionadas com as outras, o que é ilógico para análises multivariadas. Lembrem-se que estas análises estão buscando por associações que se repetem entre variáveis e que Gauch (1982a) chamou de "associações de variação coordenada". Variáveis que carregam a mesma informação são, em algum sentido, redundantes e daí vem o termo "análise de redundância" empregado para denominar algumas técnicas multivariadas. São essas associações que formam o padrão nos dados. Se selecionarmos as variáveis de modo a reduzir as associações, os principais padrões ficarão fora de nossos dados, e as análises só poderão captar padrões secundários. Por outro lado, freqüentemente algumas variáveis representam a mesma coisa medida de formas diferentes. Este problema é chamado "*aliasing*" (Mac Nally 1994) e dependendo de sua associação, os "*aliases*" podem mascarar ou realçar padrões. Considere a questão a respeito dos principais padrões nas características de riachos. Se incluirmos concentração de cálcio, pH, condutividade elétrica e profundidade, o maior padrão, e portanto a dimensão mais importante (ou "eixo", ou "variável fantasma") seria associada com a química das águas. Se incluirmos um grande número de variáveis que representam as dimensões físicas do riacho, como largura, área da seção transversal, descarga, distância do mar e área do espelho livre de vegetação, poderíamos mudar o maior eixo para um que representasse variáveis físicas. Em outras palavras, dependendo das variáveis que se escolha, pode-se criar qualquer padrão que se queira. A seleção de variáveis dependentes deve ser feita segundo critérios que sejam independentes de qualquer variável externa (independente) que será examinada, ou as inferências obtidas não terão significado.

[Muitos índices ecológicos podem ser matematicamente tratados como se fossem univariados. Entretanto, eles são essencialmente multivariados (compostos de muitas variáveis dependentes) e sua interpretação é complexa, da mesma forma que qualquer estatística multivariada.] Os exemplos incluem biomassa, riqueza de espécies, diversidade e produtividade. Todos estes são gerados por um número de outras variáveis e considerá-las como unidades podem levar-nos a conclusões ilógicas (Magnusson 2002). Por exemplo, dois locais podem ter riquezas de mamíferos idênticas. Uma possui apenas espécies de roedores silvestres, enquanto outra possui roedores silvestres, ungulados, carnívoros e primatas. Poucas pessoas considerariam-nas como sendo igualmente diversas.

Alternativamente, o objetivo de um estudo poderia ser o aumento da diversidade de mamíferos em uma reserva. Na reserva ocorrem 100 antas, 100 queixadas, 50 capivaras, 50 pacas, 50 cutias e duas onças pintadas. As onças reduzem mais as densidades de cutias, pacas e capivaras do que de presas mais difíceis como antas e queixadas. Sem as onças, poderia haver 60 capivaras, 60 pacas e 60 cutias, e as densidades de antas e queixadas se reduziriam para 85 cada, tanto em termos absolutos (via competição) ou estatisticamente (porque suas densidades relativas cairíam, devido ao aumento da densidade de outras presas, um efeito que seria percebido, mesmo que o número absoluto de antas e queixadas permanecesse constante). Portanto, fica claro que a maneira mais econômica de aumentar a diversidade de mamíferos na reserva é caçar as onças. A reserva originalmente possui uma diversidade medida pelo índice "H" de Shannon igual a 1,58. Com a perda de uma espécie e dois indivíduos, o índice passa para H = 1,59, marginalmente maior. A perda de uma espécie é compensada pelo aumento da equitabilidade das espécies remanescentes. É claro que isto não é o que passa na cabeça da maioria das pessoas quando estão falando em aumentar a diversidade biológica.

Estejam alertas quanto a interpretação simples de estatísticas multivariadas, mesmo quando vêm mascaradas como se fossem medidas univariadas. Assim como para outras estatísticas multivariadas, precisamos estar certos do que elas representam em termos das variáveis originais, para podermos usá-las com sabedoria.

Saber o que queremos antes de começar

Isto termina nosso passeio breve e superficial nas técnicas multivariadas. Os gráficos, diagramas e valores de probabilidade associados com análises multivariadas são fáceis de ser gerados nos modernos programas de computadores. Entretanto, esperamos que este breve apanhado tenha mostrado que a interpretação destes resultados é difícil e dependente do exame dos dados originais. Algumas vezes, podemos responder questões apenas olhando as tabelas e os gráficos originais, se forem organizados de forma apropriada (Ehrenberg 1981). Dedidir se queremos saber se variáveis externas causam algum padrão ou causam os principais padrões nos dados é muito importante (Økland, 1996). Os pesquisadores parecem entender isto para as questões univariadas. Por exemplo, qualquer um reconhece a diferença entre as seguintes afirmativas: (1) os predadores são a causa dos padrões cíclicos observados em determinadas populações de roedores. Isto estabelece que os

predadores causam o maior padrão. (2) Os predadores têm algum efeito sobre o padrão de flutuação em número da população dos roedores. Isto estabelece que os predadores causam algum padrão, mas não necessariamente o mais conspícuo. Entretanto, poucos pesquisadores parecem entender esta diferença em análises multivariadas.

Não é uma questão de gerar hipóteses em oposição a testar hipóteses. Se um pesquisador precisa de um computador para gerar hipóteses por si, não deveria estar fazendo ciência. Uma classe de questões requer análise de gradientes indiretos para detectar os principais padrões antes de se testar qualquer coisa. Isto pode ser considerado como uma suavização multivariada dos dados (DeAth 1999). Uma outra classe de questões envolve análise de gradientes indiretos e não se preocupa se os padrões que emergem são ou não os principais naquele sistema biológico (Økland, 1996). Muitas questões multivariadas podem ser respondidas simplesmente por uma matriz de gráficos de dispersão (cada variável exibida contra as demais) (Basford e Tukey 1999). Antes de aplicar estatísticas mais complicadas, certifique-se de que pode criar seus próprios gradientes hipotéticos. Comece com programas como COMPAS (Minchin 1987) para ter certeza de que compreendeu os fatores básicos que criam (ou escondem) padrões nos dados.

Inferências carregadas de significado freqüentemente podem ser feitas sem o uso de técnicas complicadas, simplesmente rearranjando a ordem dos atributos e objetos nas tabelas (Braun-Blanquet 1932, Gauch 1982). Exercitem as técnicas em dados simulados para se certificar que elas detectam os padrões nos quais estão interessados. Se acharem que construir gradientes é uma tarefa difícil, provavelmente ainda não estão preparados para encontrá-los. A maioria dos "especialistas" em análises multivariadas na realidade são especialistas no uso de um ou outro programa. A arte do uso de análises multivariádas está na escolha das opções mais apropriadas para detectar os padrões esperados nos seus dados, e as opções mais apropriadas podem não estar no programa do seu computador. Nós levamos muitos anos para começar fazer opções menos ruins em análises multivariadas, e recomendamos que os leitores exercitem cautela durante o tempo necessário para apreender as limitações destas análises antes de tentar explorar suas grandes potencialidades. Passada esta etapa, voce estará apto a fazer uso dos dons divinos de Pandora, sem se contaminar muito com o conteúdo de sua caixa.

CAPÍTULO 13

> [O pesquisador só é bem sucedido quando consegue explicar seus resultados para os não pesquisadores]

Como escrever melhor de trás para frente

Muitos estudantes passam a maior parte de seu tempo preocupados com testes estatísticos, mas escrever o manuscrito normalmente é a parte mais difícil. É claro que será mais fácil se o pesquisador souber, desde o começo, que a coleta de dados e os tratamentos estatísticos são parte do processo da comunicação e que os erros cometidos nas fases iniciais do estudo comprometem a comunicação oral e escrita dos resultados, ao final do trabalho.

Entretanto, temos visto que há vícios culturais poderosos que atuam dentro da cultura acadêmica, e que nos induzem a coletar dados que não servem para responder a questão formulada, mas que são eficientes para propiciar identificação cultural. É só quando tentamos comunicar resultados de pesquisa para outras pessoas, especialmente quando tentamos comunicar com pessoas que não fazem parte da cultura científica, que descobrimos se fomos bem sucedidos ao coletarmos e analisarmos os dados de nosso estudo.

Escrever não é fácil, e o pior é que não melhora muito com a experiência. Pesquisadores competentes não ficam repetindo o mesmo estudo, para cada vez fazê-lo melhor. Depois de dominar um aspecto, eles se dedicam a investigar e comunicar sobre um fenômeno ainda mais complexo. As dificuldades com a comunicação aumentam a cada novo estudo, especialmente se o pesquisador tiver de lançar mão de manipulações estatísticas complexas, com as quais poucos leitores estarão familiarizados. Basicamente, nossos fluxogramas vão ficando mais complicados à medida que aprendemos sobre o mundo.

Entretanto, há algumas regras simples que podemos usar para nos mantermos no caminho de uma comunicação cristalina rumo de comunicação efetiva. Magnusson (1996) chamou este processo de "como escrever de trás para frente" (how to write backwards).

As questões científicas geralmente podem ser divididas em dois grupos, baseado em sua complexidade. Questões de

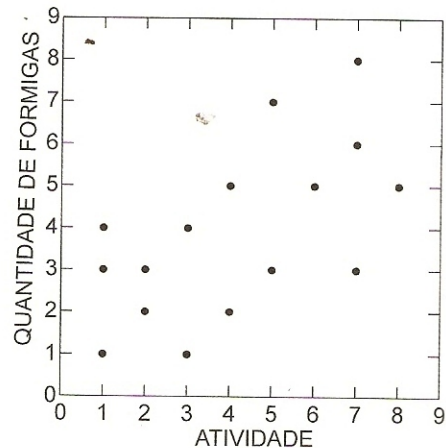
"como ?" e questões de "o quê ?" são geralmente simples, e usualmente estão associadas ao uso trivial das estatísticas descritas nas primeiras seções deste livro. Perguntas do tipo "por quê?" são quase sempre mais complexas e há um grande debate na literatura científica a respeito de como formular hipóteses nulas apropriadas para questões deste tipo (veja Connor e Simberloff 1986).

Um Esquema Conceitual Simples

Começaremos nossa discussão com uma pergunta simples do tipo "como ?", já que isto facilitará nossa comunicação. Entretanto, questões do tipo "como ?" e "o quê ?" usualmente não interessam muitas pessoas, a não ser que a resposta seja muito inesperada. Por exemplo, durante séculos, ninguém se preocupou em perguntar "como os filhotes de crocodilos conseguem alcançar a segurança das águas após eclodirem de seus ovos?". Os ovos de crocodilos são enterrados em covas que a fêmea abre nas margens dos corpos distantes d'água onde vivem. Quando Tony Pooley mostrou que as enormes e ferozes mães desenterram os ovos, ajudam gentilmente os filhotes a romperem as cascas e carregam-nos em sua boca até a água, todas as pessoas acharam a questão muito interessante. Vamos ilustrar a técnica de escrever de trás para frente usando uma questão do tipo "como ?", e , ao final deste capítulo, brevemente vamos discutir sobre as dificuldades de se enfrentar questões do tipo "por quê ?".

[O primeiro passo no processo é escrever quais as principais conclusões do estudo.] Elas devem ser escritas de forma simples e sem modificadores. Por exemplo, podemos ter concluído que "a quantidade de formigas na dieta de alguns anuros é relacionada com a intensidade de forrageio". Tendo em mente que as relações podem estar ocultas ou mesmo distorcidas por alguma(s) outra(s) variável(is), precisamos ver se podemos produzir gráficos bi-dimensionais que ilustrem a maneira como a quantidade de formigas na dieta varia com a intensidade de forrageamento, sem o efeito de variáveis que possam causar confusão. A quantidade de formigas na dieta dos anuros varia em função do seu tamanho e das estações, mas para simplificar vamos assumir que o pesquisador usou técnicas estatísticas apropriadas para isolar o efeito da atividade de forrageamento dos efeitos da estação e do tamanho, e foi capaz de produzir um gráfico das parciais em duas dimensões que ilustra a conclusão (Figura 71).

FIGURA 71

[Somente precisamos apresentar os resultados que nos levam a conclusão.] Para avaliar o quanto o gráfico dá suporte à conclusão, primeiro temos de estar certos de que as dimensões do gráfico refletem as dimensões da conclusão. Nossa conclusão é que a quantidade de formigas varia com a intensidade de

forrageamento. Portanto, o eixo Y (variável dependente) deve representar a "quantidade de formigas" e o eixo X (variável independente) deve representar a "intensidade de forrageamento". Não há alternativas.

Já que os eixos devem representar estas dimensões, e somente estas dimensões, os gráficos realmente úteis são usualmente gráficos de parciais (o efeito da variável de interesse, depois que os efeitos de todas as outras variáveis relevantes tiverem sido controlados experimental ou estatisticamente).

[Se os eixos do gráfico representarem qualquer coisa que não seja as dimensões de nossa conclusão, significa que não temos evidência convincente de nossa conclusão.]

Se tivermos feito manipulações estatísticas, pode ser necessário informar sobre valores de probabilidades associados com a hipótese nula que obtivemos nos testes. Vimos que testes estatísticos freqüentemente podem ser substituídos por gráficos simples e muitos autores afirmam que testes estatísticos nunca deveriam ser usados. Entretanto, também vimos que manipulações de dados podem fazer com que padrões fracos pareçam mais fortes do que eram nos dados originais e que algumas técnicas, como função discriminante e análise de correspondência canônica, podem produzir gráficos que podem nos enganar a respeito dos fatores afetando nossos organismos em estudo. Estes gráficos não deveriam ser apresentados sem uma justificação probabilística.

A figura 71 é convincente. Podemos ver que há uma relação forte entre a atividade de forrageamento e a quantidade de formigas na dieta. Entretanto, em muitos estudos, encontramos gráficos que não suportam a conclusão, e temos de modificar a conclusão antes de continuar. Uma vez que tivermos o gráfico correto, podemos redigir a seção de métodos. [A seção de métodos deve apresentar toda a informação necessária para mostrar como obtivemos os dados para nossos gráficos, e nada mais.] Geralmente, não é necessário informar que as medições foram feitas em um recinto chamado laboratório, ou que pesamos os organismos em uma balança analítica, se a massa dos organismos não aparece nas análises. A idéia é escrever com o mínimo de palavras possíveis, de forma que o leitor possa se concentrar nos aspectos importantes, e isto vai maximizar a comunicação. Algumas vezes, teremos de modificar nossa conclusão neste ponto e começar tudo de novo, porque reconhecemos alguma falha fatal no protocolo de obtenção dos dados. Bons pesquisadores usualmente não se importam em reconhecer quando cometeram erros; isto mostra que aprenderam alguma coisa.

Na seção de métodos, temos que explicar também porque acreditamos que cada ponto no gráfico representa uma informação independente em relação a nossa questão (em outras palavras, como evitamos cair em pseudo-repetição). Se tivermos mais pontos no gráfico do que observações independentes, o gráfico será enganoso. Neste caso, teremos que refazer o gráfico e reavaliar nossas conclusões.

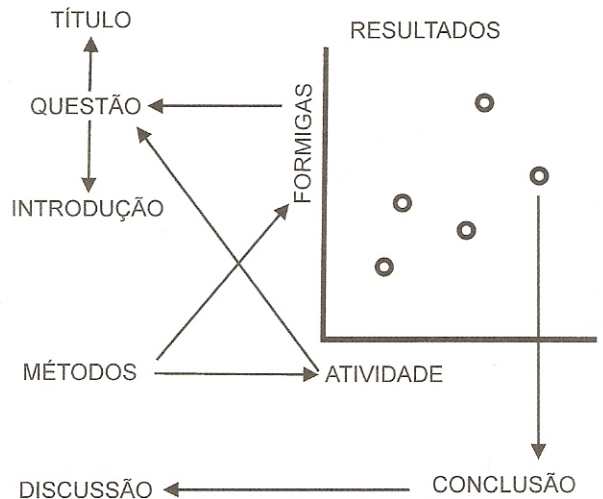
Tendo definido nossa conclusão e estando confiantes de que nossos resultados realmente apóiam aquela conclusão, estamos prontos para redigir a introdução. [A introdução deve conter o mínimo de informação necessária para apresentar a questão para a qual nossa conclusão é a resposta.] A questão pode ter sido levantada baseado na literatura científica,

informação de pesquisadores experientes ou a partir de observações preliminares. Não precisamos de mais nada. Revisões de literatura a respeito da taxonomia da(s) espécie(s) envolvida(s), ou professar nossa fé na necessidade da conservação das florestas tropicais, ou qualquer outro discurso "científico", só será relevante se estes assuntos introduzem os nomes de dos eixos de nossos gráficos.

Após termos definido nossas questões e respostas, podemos dar início à redação da "discussão". [A discussão deve apresentar somente a informação (usualmente obtida da literatura científica) que apóia, estende, modifica ou contraria as conclusões que se baseiam nos dados que apresentamos.] Não é um fórum para apresentação de conclusões que não são baseadas nos dados do presente estudo. Quando lemos artigos de outros pesquisadores, não queremos mergulhar em páginas e páginas de especulações, tentando filtrar o que realmente é relevante no estudo Portanto, faça pelos outros o que gostaria que fizessem por você.

A última coisa a se preocupar é com o título. [Os melhores títulos usualmente são questões ou respostas (conclusões).] Portanto, podemos usar a questão "A quantidade de formigas na dieta de anuros varia com a intensidade de forrageio ?", ou a resposta "A quantidade de formigas na dieta de anuros varia com a intensidade de forrageio". Há diferentes maneiras de expressar estas idéias, mas o ponto crítico é que as dimensões na conclusão (e na questão e no gráfico) devem estar explícitas no título. De outra forma, os leitores potenciais não saberão o que esperar e talvez não se sentirão motivados a ler o texto.

Se você está tendo problemas em expressar suas idéias, verifique se as seções de seu texto estão ligadas (figura 72). Este esquema simples é surpreendentemente eficaz em identificar problemas de comunicação. Não importa quão experiente você seja, um fluxograma conceitual (como o da figura 72), será sempre necessário ao redigir novos artigos. Se você tem muitas conclusões, é porque provavelmente não descobriu nada muito excitante, ou porque confundiu resultados metodológicos intermediários com os resultados que realmente afetam sua conclusão biológica. Mesmo artigos científicos muito complexos raramente têm mais do que duas ou três conclusões principais.

FIGURA 72

Detalhes aborrecidos, mas importantes

Há muitas outras complicações, especialmente quando escrevendo trabalhos do tipo "por quê". Usamos o exemplo dos anuros para introduzir algumas das complicações com as quais os autores iniciantes geralmente acabam se deparando, mas problemas similares surgem não apenas na ecologia animal, mas em todos os campos da ciência.

Um erro comum é assumir que nossas medições correspondem a quantidades conceituais reconhecidas por outros pesquisadores. Todas as unidades de medidas são índices de alguma quantidade teórica, física ou conceitual. Por exemplo, para medir nossos sapos usamos uma régua ou um paquímetro marcado com distâncias padronizadas. Estas distâncias padronizadas guardam uma relação aproximada com uma barra de metal guardada perto de Paris, que é o padrão internacional para um metro. Estamos interessados em tamanhos de anuros, e muitos pesquisadores usam o comprimento do corpo como um índice de tamanho, mas a definição de "tamanho" no dicionário é "corpo, dimensão, volume". Sob a definição do dicionário "volume", que é a única que define qual a dimensão que "tamanho" representa, a massa provavelmente seria um índice melhor de tamanho do que o comprimento. Um conceito tão simples quanto "tamanho" pode significar coisas diferentes para diferentes pessoas.

Nosso gráfico tem "quantidade de formigas" no eixo Y, mas não definimos "quantidade". Alguns pesquisadores usam o número de presas para representar "quantidade", alguns usam o volume de presas, outros usam o conteúdo calorífico das presas, e muitos outros usam diferentes índices que podem ser encontrados na literatura. Nenhum desses é necessariamente superior aos outros. Na verdade, o melhor índice dependerá da questão que está sendo formulada. Os problemas surgem quando o pesquisador usa um determinado índice porque é o índice freqüentemente usado por outros pesquisadores (uma norma cultural), e não porque é o mais apropriado para a questão formulada. Por exemplo, para estudar a dieta de animais insetívoros, como anuros, muitos pesquisadores usam duas medidas lineares como indicativo do volume das presas. A estimativa resultante tem pouca relação com o volume medido por deslocamento de fluido (Magnusson et al. 2003). Esses pesquisadores estão usando um índice de volume que não tem contraparte equivalente no mundo físico, portanto, o potencial para problemas de comunicação é grande. "Intensidade de forrageio" é ainda mais complicado. Pesquisadores têm usado a freqüência de movimento, a distância movida por unidade de tempo, a área coberta por unidade de tempo, a proporção de tempo que os organismos permaneceram em movimento, ou combinações multivariadas destas coisas, ou ainda muitas outras medidas para indexar a intensidade de forrageio. Cada um destes índices nos diz coisas diferentes a respeito da atividade de anuros.

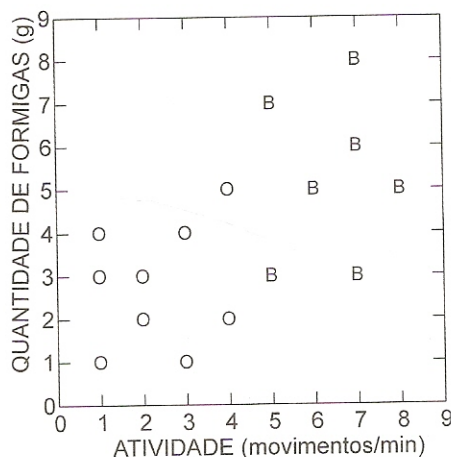
> A escolha certa de um índice só pode ser feita depois que se decidida qual a questão que se quer responder.

Usar um índice padrão por que ele permite comparações entre estudos faz sentido apenas se estes estudos investigaram exatamente a mesma questão. Ao invés de mergulhar em um estudo do tipo "eu também", pode ser mais

interessante levantar uma questão que outros pesquisadores ainda não tenham respondido. A escolha de um índice (medida, variável) inapropriado no início do estudo pode nos limitar a atacar apenas questões que já foram respondidas por outros, e não guardam mais grande interesse. Os pesquisadores mais sagazes usam um esquema como o da figura 72, ou alguma outra coisa, junto com um gráfico hipotético para dar início à redação de seu artigo, antes mesmo de começar a coletar os dados.

Questões do tipo "Por quê?"

Determinamos que há uma relação entre a atividade de forrageio e o número de formigas na dieta. Isto foi fácil, mas o processo se complica quando perguntamos "Por quê há uma relação entre a atividade de forrageio e a quantidade de formigas na dieta?". Os bufonídeos (serão assinalados como "B" na Figura 73) tendem a ser a família de anuros mais ativa, e outras famílias (serão assinaladas como "O" na Figura 73) geralmente são menos ativas. Talvez tenha sido apenas coincidência de os ancestrais dos bufonídeos serem ativos e comerem formigas. Se não estivermos interessados nos efeitos da inércia filogenética, poderíamos usar contrastes filogeneticamente independentes (Garland et al. 1992) para remover os efeitos de parentesco evolutivo entre as espécies, ou talvez pudéssemos investigar o quanto a dieta e o forrageio são relacionados dentro de cada espécie, usando indivíduos como as unidades amostrais. Dependendo de nossa questão, os efeitos filogenéticos podem ser o foco do estudo ou simplesmente uma variável incômoda, que precisa ter seus efeitos controlados.

FIGURA 73

Muitas hipóteses têm sido propostas para explicar a relação de quantidade de formigas na dieta e o comportamento de forrageio nos anuros. Pode ser que anuros ativos tendem a encontrar mais formigas porque eles tendem a se locomover mais. Anuros que comem formigas podem seqüestrar veneno de suas presas, tornando-se impalatáveis para seus próprios predadores, tornando-os menos vulneráveis enquanto passeiam por aí. As formigas podem representar recursos concentrados que só podem ser explotados por predadores que se locomovem muito. Talvez os anuros só possam ser continuamente ativos se usam uma fonte de alimento com pouca tendência a flutuar espacial e temporalmente. Algumas destas hipóteses indicam que as diferenças nas dietas são reflexos passivos do comportamento de forrageio, e outras indicam que o comportamento de forrageio evoluiu para maximizar o encontro com formigas. A situação é ainda mais complicada, porque o que determina a variação da quantidade de formigas dentro de cada espécie pode não ser o que determina a variação na quantidade de formigas entre espécies.

Não temos a resposta para esta pergunta do tipo "Por quê?". Muitos pesquisadores estão desenvolvendo experimentos sofisticados com rigorosos esquemas de amostragem para tentar se aproximar da resposta. Talvez alguém vá responder a pergunta, ou algum aspecto da pergunta, ou talvez não. Mas, podemos estar certos de que uma resposta inequívoca não virá de algum teste estatístico padrão tirado de um manual de métodos para ecólogos. Podemos estar certos

também que nossa confiança na conclusão tem pouco a ver com a probabilidade de encontrar uma linha de regressão com uma inclinação igual ou maior que a esperada se a relação entre a quantidade de formigas na dieta e a atividade de forrageio fosse somente devida ao acaso.

[A ciência avança vagarosamente com a acumulação de respostas para perguntas dos tipos "como?" e "o quê?" e avança rapidamente com respostas para perguntas do tipo "por quê?".] Os grandes cientistas e filósofos, como Fisher e Popper, estavam interessados em perguntar "por quê?". É uma pena que seus nomes estejam associados com simples técnicas que são mais apropriadas para responder perguntas de "como?" e "o quê?". Descobrir como responder uma pergunta do tipo "por quê?" é um ato de genialidade. Deixamos vocês com esta tarefa e esperamos que nossa "receita" de como escrever de trás para frente possa ajudá-los a comunicar suas conclusões para as outras pessoas.

Uma palavra final

Este livro termina aqui. Esperamos que ele tenha contribuído para aumentar sua capacidade de entender a literatura científica e de se fazer entender por seus colegas. Como dissemos anteriormente, a habilidade de delinear bons experimentos científicos é uma arte, e não pode ser ensinada em livros. Entretanto, esperamos ter apontado algumas "armadilhas" freqüentes no delineamento amostral, especialmente aquelas que têm origem na bagagem cultural e que facilmente passariam despercebidas se alguém não nos tivesse alertado sobre elas. A "academia" passou anos a fio inculcando em vocês a mensagem de que o mais importante em suas vidas profissionais era parecer um "cientista", se vestindo, falando e escrevendo da forma complicada como fazem os "cientistas". Entretanto, se quiserem comunicar resultados de pesquisa, vão precisar ir além de comunicar apenas identidade cultural. Se depois de lerem este livro sentirem-se tentados a apresentar seus resultados em alguns gráficos simples, e a usarem a estatística para simplificar problemas de muitas dimensões, então nossos esforços não foram em vão.

ÍNDICE REMISSIVO DE ASSUNTOS

Agrupamento de mensagem de menor comprimento "Minimum message length clustering", 120
Amplitude, 26
Análise
de Agrupamento, 118
de Coordenadas Principais PCoA, 111
de Caminhos, 91, 92
de Componentes Principais PCA, 111
de Correspondência CA, 111
de Correspondência Canônica CCA, 115
de Correspondência Destendenciada "Detrended correspondence analysis DCA", 111
de Covariância ANCOVA, 78
Cladistas, 118
de Curvas Principais "Principal Curves Analysis", 112
direta de Gradiente, 104
de Função Discriminante DFA, 117
indireta de Gradiente, 109
não-métrica de escalas multidimensionais NMDS, 109
Path Analysis, 15, 91
de Redundância Canônica CRDA, 113
de Redundância RDA, 113
de Resíduos, 78
de Sistemas, 15
de Variância, 58
de Variância Multivariada MANOVA, 114
ANOVA, 58
Árvores de regressão, 120
Associação de Gower, 113
Atributos, 107
Bancos de dados, 5
Biomassa, 121
Bonferroni, 56
Bootstrap, 120
Bray-Curtis, 113
Caminhos "path", 91
CCA não lineares, 115
Chaves dicotômicas, 118
City-block, 107
Cluster analysis, 118
Coeficiente de correlação de Pearson, 111
Coeficiente de regressão, 69
Coleta de dados substitutos, 34
Composição de espécies, 107
Correlação, 18
Canônica, 103
Cultura, 15
Curva teórica, 51
Dados, 25
exploratórios, 87
DECORANA, 111
Delineamento amostral, 5, 7
Dendrograma, 118
divisivo, 118
aglomerativo, 118
Descritores de dispersão, 26
Desvio
médio absoluto, 26
padrão, 26
Dimensão, 107, 109
Dissimilaridade, 107, 108
Distância 107
"Manhattan", 107
euclidiana, 112
Distribuição normal, 28, 52
Diversidade, 121
Dummy variables, 114
Efeitos
diretos, 91
gerais, 91
indiretos, 19, 92
Eixos, 109
Elevação, 68
Erro
estocástico, 95
padrão, 28
tipo I, 54
tipo II, 54
Escala, 22
adimensional, 106

de inferência, 38
Escopo, 38
Espaço multidimensional, 116
Estatística, 27
Bayesiana, 20
F, 57
freqüentista, 42
inferencial, 15
multivariada, 103
popperiana, 42
t, 48
Estimativa
não-linear, 97
padronizada, 91
Experimento, 20
Exploração, 116
Fator, 60, 79
fixo, 65
randômico, 65, 79
Filogenia, 118
Filosofia, 24
Fisher, Ronald 56, 57
Fluxograma, 15
Função
de ligação, 99
de perda "loss function", 98
Gradiente
análise direta de, 104
ecológico, 108
análise indireta de, 107
Gráficos, 24
de dispersão, 31
parciais, 76
Graus de liberdade, 49
Hipótese, 16, 42
nula, 42
Hotelling-Lawley Trace, 103
Inclinação, 68, 75
Índice de diversidade "H", 121
Inferência forte, 20
Informação, 33
Inteligência artificial, 82
Interações, 78
Intervalos de confiança, 29
Lambda (8) de Wilks, 113
Maior eixo reduzido, 69
Matriz
de associação, 107
de similaridade, 107
Média, 26
dos quadrados MQ, 61
Medida
geométrica, 112
de associação, 107
de similaridade, 107
Método científico, 41
Modelos, 17
de equação estrutural, 91
gerais lineares, 99
Nível de significância, 53
NMDS, 107
Números aleatórios, 82
Objetos, 107
Observações independentes, 34
Ordenação, 104
direta, 104
Outlier, 58
Parâmetros, 27, 48
Pillai Trace, 103
Popper
Karl, 41
Filosofia de , 20
Populações
biológicas, 28
estatísticas, 27
Probabilidade, 20, 42
Problemas inversos, 94
Processo de decisão, 16
Procrustes, sobreposição de, 115
Produtividade, 121

Projetos
- integrados, 05
- interdisciplinares, 05
- multidisciplinares, 05

Pseudoprobabilidade, 84
Pseudo-repetição, 34
- espacial, 35
- filogenética, 36
- técnica, 36
- temporal, 35

Qui-quadrado, 111
Razão da variação, 57
Regressão
- logística, 99
- do maior eixo reduzido (média geométrica), 98
- do maior eixo, 68
- dos mínimos quadrados, 69
- múltipla, 76
- parcial, 76
- polinomial, 115

Relações
- causais, 18
- filogenéticas, 118
- lineares, 21, 89

Repetições, 34
Similaridades, 108
Sinapomorfismos, 118
Soma de quadrados *SQ*, 61, 69
Step-down, 44
Step-up, 82
Step-wise, 82
Subgrupo, 118
Tabelas
- de contingência, 37
- de valores críticos, 49

Técnicas freqüentistas, 20
Teorema do Limite Central, 28
Teste
- estatístico, 20
- de Levene, 58
- de Mantel, 114
- não-paramétrico, 114
- de Tukey, 59

Transformação, 71, 95
Tratamento, 61
Unidades amostrais, 9, 38
- forma e tamanho, 9

Universo
- amostral, 38
- de inferência, 38
- de interesse, 38

Validação, 87, 116
Valor
- esperado, 62
- crítico, 49

Variabilidade, 27
- residual, 60

Variância, 27
- composta, 79
- homogeneidade das, 58, 64
- partição das, 60

Variável
- aleatória, 84
- categórica, 79
- contínua, 79, 109
- dependente, 19
- independente, 19

Verossimilhança máxima, 99, 100
Vetores "eigen", 110

REFERÊNCIAS BIBLIOGRÁFICAS

Abuabara, M. A. P. e M. Petrere. 1997. Estimativas da Abundância de Populações Animais. EDUEM, Maringá.

Albert, J. 1997. Teaching Bayes' rule: a data oriented approach. *American Statistician* 51:247-253.

Allen, T. F. H e T. B. Starr, 1982. Hierarchy: perspectives for ecological complexity. University of Chicago Press, Chicago.

Allen, T. F. H. 1998. The landscape "level" is dead: persuading the family to take it off the respirator. Pp. 35-54 In Peterson, D. L. e Parker, V. T. (eds) Ecological Scale. Colombia University Press, New York.

Anderson, D. R., K. P. Burnham, G. C. White, e D. L. Otis. 1983. Density estimation of small-mammal populations using a trapping web and distance sampling methods. *Ecology* 64:674-680.

Anderson, D. R., K. P. Burnham, W. R. Gould e S. Cherry. 2001. Concerns about finding effects that are actually spurious. *Wildlife Society Bulletin* 29(1):311- 316.

Anderson, D. R.; Burnham, K. P.; Thompson,W. L. 2000. Null hypothesis testing: problems, prevalence, and an alternative. *Journal of Wildlife Management* 64:912-923.

Anderson, M. e P. Legendre. 1999. An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *Journal of Statistical and Computer Simulation* 62:271-303.

Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26:32-46.

Anderson. M. J. & T. J. Willis. 2003. Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology* 84(2):511-525.

Anscombe, F. J. 1973. Graphs in statistical analysis. *American Statistician* 27:17- 21.

Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* 157:101- 118.

Azevedo-Ramos, C.,W. E. Magnusson e P. Bayliss. 1999. Predation as a key factor structuring tadpole assemblages in a savanna area in central Amazonia. *Copeia* 1999:22-33.

Bard, Y. 1974. Nonlinear Parameter Estimation. Academic Press, New York.

Basford, K. E. e J.W. Tukey. 1999. Graphical analysis of multiresponse data illustrated with a plant breeding trial: interdisciplinary statistics. Chapman and Hall, Boca Raton.

Belbin, L. 1992. PATN Pattern Analysis Package – Technical Reference. CSIRO, Canberra.

Benedetti-Cecchi, L. 2003. The importance of the variance around the mean effect size on ecological processes. *Ecology* 84:2335-2346.

Benjamini, Y. e Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 57 (Series B): 289-300.

Bennington, C. C. e W. V. Thayne. 1994. Use and misuse of mixed model analysis of variance in ecological studies. *Ecology* 75:717-722.

Berk, K. N. 1978. Comparing subset selection procedures. *Technometrics* 20:1-6.

Beyer, W. H. 1968. Handbook of Probability and Statistics. CRC Press, Boca Raton, Florida, USA.

Borg, I. e P. Groenen. 1997. Modern Multidimensional Scaling: Theory and Applications. Springer, New York, USA 135

Bradshaw, G. A. 1998. Defining ecologically relevant change in the process of scaling up: implications for monitoring at the landscape level. Pp. 227-249 In Peterson, D. L. e Parker, V. T. (eds) Ecological Scale. Colombia University Press, New York.

Braun-Blanquet, J. 1932. Plant Sociology: The Study of Plant Communities. Hafner, London.

Brown, I. F. 1997. Let the people judge: community examining boards help students learn to communicate. *Bulletin of the Ecological Society of America* 78:210- 211.

Burnham, K. P.; Anderson, D. R. 1998. Model selection and inference: a practical information-theoretic approach. Springer-Verlag, New York.

Cade, B. S., J. W. Terrell e R. L. Schroeder. 1999. Estimating regression effects of limiting factors with regression quantiles. *Ecology* 80:311-323.

Callaghan, A. e G. J. Holloway. 1999. The relationship between environmental stress and variance. *Ecological Applications* 9:456-462.

Carlson, J. M. e J. Doyle. 1999. Highly optimized toelrance: a mechanism for power laws in designed systems. *Physical Review* E 60:1412-1427.

Carpenter, S. R. 1999. Microcosm experiments have limited relevance for community and ecosystem ecology: a reply. *Ecology* 80:1085-1088.

Caughley, G. e A. R. E. Sinclair. 1994. Wildlife Ecology and Management. Blackwell Scientific Publications, Oxford.

Cherry, S. 1999. Statistical tests in publications of The Wildlife Society. Wildlife Society Bulletin 26: 947-953.

Connolly, J., P. Wayne e F. A. Bazzaz. 2001. Interspecific competition in plants: how well do current methods answer fundamental questions. *American Naturalist* 157:107-125.

Connor, E.F. e D. Simberloff. 1986. Competition, Scientific Method and null models in ecology. *American Scientist* 74:155

Dale, M. B. 1988. Knowing when to stop: cluster concept – concept cluster. *Coenoses* 3:11-32.

Day, R.. W. e G. P. Quinn. 1989. Comparisons of treatments after an analysis of variance in ecology. *Ecological Monographs* 59:433-463.

De'Ath, G. 1999. Principal curves: a new technique for indirect and direct gradient analysis. *Ecology* 80:2237-2253.

Deming, W. E. 1975. On probability as a basis for action. *American Statistician* 29:146-152.

Dytham, C. 1999. Choosing and using statisitics: a biologist's guide. Blackwell Science, Oxford.

Ehrenberg, A. S. C. 1981. The problem of numeracy. *American Statistician* 35:67- 71.

Ellison, A. M. 1993. Exploratory data analysis and graphic display. Pp 14-45 In Scheiner, S. M. e J. Gurevitch (eds) Design and Analysis of Ecological

expectations. *American Naturalist* 135:310-317.

Experiments. Chapman and Hall, New York.

Faith, D. P., P. R. Minchin e L. Belbin 1987. Compositional dissimilarity as a robust measure of ecological distance: a theoretical model and computer simulations. *Vegetatio* 69:57-68.

Feinsinger, P. 2002. Designing field studies for biodiversity conservation. Island Press,Washington, USA.

Fortin, M. J. e J. Gurevitch 1993. Mantel tests: spatial structure in field experiments. Pp 342- 352 In Scheiner, S. M. e J. Gurevitch (eds) Design and Analysis of Ecological Experiments. Chapman and Hall, New York.

Fowler, N. 1990. The 10 most common statistical errors. *Bulletin of the Ecological Society of America* 71:161-164.

Freedman, D. A. 1983. A note on screening regression equations. *American Statistician 37*:152- 155.

Friendly, M. 1995. Conceptual and visual models for categorical data. *American Statistician 49*:153-160.

Gaines, S. D. e W. R. Rice. 1990. Analysis of biological data when there are ordered

Garland, T., P. H. Harvey e A. R. Ives. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic Biology* 41:18-32.

Gauch, H. G. 1982a. Multivariate analysis in community ecology. Cambridge University Press, Cambridge.

Gauch, H. G. 1982b. Noise reduction by eigenvector ordinations. *Ecology* 63:1643-1649.
Goleman, D. 1995. Emotional Intelligence. Bloomsbury Publishing, London.
Green, R. H. 1989. Power analysis and statistical strategies for environmental monitoring. *Environmental Research* 50:195-205.
Guisan, A., T. C. Edwards e T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling* 157:89-100.
Guttman, L. 1985. The illogic of statistical inference for cumulative science. *Applied stochastic models and data analysis* 1:3-10.
Hairston, N. G. 1989. Hard choices in ecological experimentation. *Herpetologica* 45:119-122.
Hale, S. S. 1999. How to manage data badly (part 1). *Bulletin of the Ecological Society of America* 80:265-268.
Hall, E. T. 1959. The silent Language. Doubleday and Company, New York. Hallgren, E., M.W.
Harris, R. J. 1975. A primer for multivariate statistics. Academic Press, New York.
Higashi, M. e T. P. Burns. 1991. Theoretical Studies of Ecosystems: The Network Perspective. Cambridge University Press, Cambridge.
Hilborn, R. e M. Mangel. 1997. The Ecological Detective. Princeton University Press, Princeton.
Hobbs, R. J. 1998. Managing ecological systems and processes. Pp. 459-484 In D. L. Peterson & V. T. Parker (eds) Ecological Scale. Colombia University Press, New York.
Huberty, C. J. 1987. On statistical testing. *Educational Researcher 16*:4-9.
Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54: 187-211.
Iglewicz, B. 1983. Robust scale estimators and confidence intervals for location. Pp 404-431 In D. C. Hoaglin, F. Mosteller e J. W. Tukey (eds). Understanding robust and exploratory data analysis. John Wiley and Sons, New York.
Jackson, D. A. 1993. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology* 74:2204-2214.
Jackson, D. A. 1997. Compositional data in community ecology: the paradigm or peril of proportions. *Ecology* 78(3):929-940.
James, F. C. e C. E. McCulloch. 1990. Multivariate analysis in ecology and systematics: Panacea or Pandora's box? *Annual Review of Ecology and Systematics* 21:129-166.
Johnson, C. R. e C. C. Field. 1993. Using fixed-effects model multivariate analysis of variance in marine biology and ecology. *Oceanography and Marine Biology Annual Review* 31:177- 221.
Johnson, D. H. 1999. The insignificance of statistical significance testing. *Journal of Wildlife Management* 63:763-772.
Kenckel, N. C. e L. Orloci. 1986. Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. *Ecology* 67:919-928.
Koele, P. 1982. Calculating power in analysis of variance. *Psychological Bulletin* 92:513-516.
Krebs, C. J. 1989. Ecological Methodology. Harper and Row, New York.
Kruskal, W. 1988. Miracles and statistics: the casual assumption of independence. *Journal of the American Statistical Association* 83:929-940.
Kuhn, T. S. 1970. The structure of scientific revolutions. 2nd ed. University of Chicago Press, Chicago.
Lawton, J. 1999. Size matters. *Oikos* 85: 19-21.
Legendre, P. 1993. Spatial autocorrelation: trouble or a new paradigm? *Ecology* 74:1659-1673.
Legendre, P. e L. Legendre. 1998. Numerical Ecology: Second English Edition. Elsevier, Amsterdam.
Legendre, P., e E. D. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129:271-280.
Legendre, P., e M. J. Anderson. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs* 69:1-28.
Lennon, J. L. 2000. Red-shifts and red herrings in geographical ecology. *Ecography* 23:101-113.
Link,W. A. 1999. Modeling patterns in collections of parameters. *Journal of Wildlife Management* 63: 1017-1027.
Luo, J. e B. J. Fox. 1996. A review of the Mantel test in dietary studies: effect of sample size and inequality of sample sizes. *Wildlife Research* 23:267-288.
Mac Nally, R. C. 1994. On characterizing foraging versatility, illustrated by using birds. *Oikos* 69:95-106.
Mac Nally, R. C. 2002. Multiple regression and inference in ecology and conservation biology: further comments on identifying important predictor variables. *Biodiversity and Conservation* 11:1397-1401.
MacCune, B. 1997. Influence of noisy environmental data on canonical correspondence analysis. *Ecology* 78:2617-2623.
Magee, B. 1976. Popper. Fontana, London.
Magnusson, W. E. 1999. Spatial independence: the importance of the question. *Wildlife Society Bulletin* 27:1112-1113.
Magnusson, W. E. 2000a. Error bars: are they the King's clothes? *Bulletin Ecological Society of America.* 81:147-150.
Magnusson, W. E. 2000b. Statistical iatrogenesis: cure it or prevent it? *Bulletin of the Ecological Society of America* 81:198-201.
Magnusson, W. E. 2001. Standard errors of survey estimates: what do they mean? *Neotropical Primates* 9:53-54.
Magnusson, W. E. 2002a. Diversity indices: multivariate candies from Pandora's box. *Bulletin of the Ecological Society of America* 83:86-87.
Magnusson, W. E. 2002b. Categorical ANOVA: Strong inference for weak data. *Bulletin of the Ecological Society of America* 83:81-86.
Magnusson, W.E. 1996. How to write backwards. *Bulletin of the Ecological Society of America* 77:88
Magnusson, W.E., A.P. Lima, W.A. da Silva e M.C. de Araújo. 2003. Use of geometric forms to estimate volume of invertebrates in ecological studies of dietary overlap. *Copeia* 2003:13-19
Magnusson,W. E. 1997. Teaching experimental design, or how to do statistics without a bikini. Bulletin of the *Ecological Society of America* 78:205-209.
Makarenkov V., e P. Legendre 2002. Nonlinear redundancy analysis and canonical correspondence analysis based on polynomial regression. *Ecology* 83:1146-1161.
Manel, S., J. M. Dias, S. T. Buxton e S. J. Ormerod. 1999. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology* 36:734-747.
Manly, B. F. J. 1997. Randomization, Bootstrap and Monte Carlo Methods in Biology. Chapman and Hall, London.
Mantel, N. A. e R.. S. Valand. 1970. A technique for nonparametric multivariate analysis. *Biometrics* 26:547-558.
McArdle, B. H. e M. J. Anderson. 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* 82:290-297.
McCune, B. e J.B. Grace. 2002. Analysis of ecological communities MJM Software Design, Gleneden Beach, Oregon
Meeks, S. L. e R. B. D'Agostino. 1983. A note on the use of confidence limits following rejection of a null hypothesis. *American Statistician* 37:134-136.
Meot, A., P. Legendre e D. Borcard. 1998. Partialling out the spatial component of ecological variation: questions and propositions in the linear modelling

framework. *Environmental and Ecological Statistics* **5**:1-27.

Miller, G. S. Olson, C. A. Parmenter, J. Pollard, E. Rexstad, T. M. Shenk, T. R. Stanley e G. C. White. 2003. Small-mammal density estimation: a field comparison of grid-based vs webbased density estimates. *Ecological Monographs* **73**:1-26.

Milligan, G. W. e M. C. Cooper. 1985. An examination of procedures for determining the number of clusters in a data set. *Psychometrika* **50**:159-179.

Minchin, P. R. 1987. Simulation of multidimensional community patterns: toward a comprehensive model. *Vegetatio* **71**:145-156.

Moore, D. S. 1997. Bayes for beginners? Some reasons to hesitate. *American Statistician* **51**: 254-261.

Mosteller, F. e J. W. Tukey. 1968. Data analysis, including statistics. Pp 80-203 In Lindzey, G. e E. Aronson (eds). Handbook of Social Psychology (2nd edition), Volume 2. Addison- Wesley, Reading Massachusetts.

Newman, J. A., J. Bergelson e A. Grafen. 1997. Blocking factors and hypothesis tests in ecology: is your statistics text wrong? *Ecology* **78**:1312-1320.

Neyman, J. 1937. Outline of a theory of statistical estimation based on the classical theory of probability. Philosophical Transactions of the Royal Society of London Series A 231:333- 380.

Noy-Meir, I., D.Walker e W. T. Williams. 1975. Data transformations in ecological ordination. II. On the meaning of data standardization. Journal of Ecology 63:779-800.

O'Neill, R. V. e A.W. King. 1998. Homage to St Michael; or why are there so many books on scale? . Pp. 3-15 In Peterson, D. L. e Parker, V. T. (eds) Ecological Scale. Colombia University Press, New York.

Økland, R. H. 1996. Are ordination and constrained ordination alternative or complementary strategies in general ecological studies? *Journal of Vegetation Science* **7**:289-292.

Økland, R. H. 1999. On the variation explained by ordination and constrained ordination axes. *Journal of Vegetation Science* **10**:131-136.

Oksanen, J. e P. R. Minchin. 1997. Instability of ordination results under changes in input data order: explanations and remedies. *Journal of Vegetation Science* **8**:447-454.

Oksanen, J. e P. R. Minchin. 2002. Continuum theory revisited: what are species responses along ecological gradients? *Journal of the Vegetation Science* **8**:447-454

Olson, C. L. 1976. On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin* **83**:579-586.

Ormerod, S. J., M. W. Pienkowsky e A. R.Watkinson. 1999. Communicating the value of ecology. *Journal of Applied Ecology* **36**:847-855.

Osenberg, C.W., O. Sarnelle, S. D. Cooper e R. D. Holt. 1999. Resolving ecological questions through meta-analysis: goals, metrics, and models. *Ecology* **80**(4):1105-1117.

Palmer e P. Milberg. 1999. Data diving with cross-validation: an investigation of broad-scale gradients in Swedish weed communities. *Journal of Ecology* **87**:1037- 1051.

Palmer, A. R. 1999. Detecting publication bias in meta-analyses: a case study of fluctuating asymmetry and sexual selection. 154:220-233.

Parmenter, R. R., T. L. Yates, D. R. Anderson, K. P. Burnham, J. L. Dunnum, A. B, Franklin, M. T. Friggens, B. C. Lubow, M. Pascual, M. e S. A. Levin. 1999. From individuals to population densities: searching for the intermediate scale of nontrivial determinism. *Ecology* **80**:2225-2236.

Pearson, K. 1901. On lines e planes of closest fit to systems of points in space. *Philosophical Magazine* **6**:559-572.

Peladeau, N. 199 6. Simstat for Windows. 5000 Adam Street, Montreal, QC, Canada, H1V 1W5. Pélissier, R., P. Couteron, S. Dray e P. Sabatier. 2003. Consistency between ordination techniques and diversity measurements: two strategies for species occurrence data. *Ecology* **84**:242-251.

Peres-Neto, P. R., e D. A. Jackson. 2001. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* **129**:169-178.

Petersen, J. E., Cornwell, J. C. e Kemp, W. M. 1999. Implicit scaling in the design of experimental aquatic ecosystems. *Oikos 85*:3-18.

Peterson, D. L. e Parker, V. T. (eds) 1998. Ecological Scale. Colombia University Press, New York. Petraitis, P. S., A. E. Dunham e P. H. Niewarowski. 1996. Inferring multiple causality: the limitations of path analysis. *Functional Ecology* **10**:421-431.

Picard, R. R. e R.D. Cook. 1984. Cross-validation of regression models. *Journal of the American Statistical Association* **79**:575-583.

Pickett, S. T. A., J. Kolasa e C. G. Jones. 1994. Ecological Understanding. Academic Press. San Diego.

Pielou, E. C. 1984. The interpretation of ecological data. Wiley, New York.

Pillar, V. D. 1999. How sharp are classifications? *Ecology 80*(8):2508-2516.

Platt, J. R. 1964. Strong inference. *Science* **146**:347-353.

Polis, G. A. 1999. Why are parts of the world green? Multiple factors control productivity and distribution of biomass. *Oikos* **86**:3-15.

Popper, K. R. 1976. Unended Quest: An Intellectual Autobiography. Fontana, London.

Powell, T. M. e J. H. Steele. 1995. Ecological Time Series. Chapman and Hall, New York.

Rao, C.R. 1964. The use and interpretation of principal component analysis in applied research. *Sankhyaa Ser.*A **26**:329-358

Rice, W. R. 1989. Analyzing tables of statistical tests. *Evolution* **43**:223-225.

Ricker, W. E. 1973. Linear regressions in fishery research. *Journal of the Fish Research Board of Canada*. **30**:409-434.

Rohlf, F.J. e R.R. Sokal. 1981. Comparing numerical taxonomic studies. *Systematic Zoology* **30**:459-490

Rohlf. F.J. 2001 Comparative methods for the analysis of continues variables: geometric interpretations. *Evolution* **55**:2143-2160

Rosenthal, R. e D. B. Rubin. 1982. A simple general purpose display of magnitude of experimental effect. *Journal of Educational Psychology* **74**:166-169.

Rosenthal, R. e D. B. Rubin. 1994. The counternull value of an effect size: a new statistic. *Psychological Science* **5**:329-334.

Salsburg, D. 2001. The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century. W. H. Freeman, New York.

Salsburg, D. S. 1985. The religion of statistics as practiced in medical journals. *American Statistician* **39**:220-223.

Sawada, M. 1999. Rookcase: an excel 97/2000 visual basic (VB) add-in for exploring global and local spatial autocorrelation. *Bulletin of the Ecological Society of America* **80**:231-234.

Shipley, B. 1999. Testing causal explanations in organismal biology: causation, correlation and structural equation modelling. *Oikos* **86**:374-382.

Short, J., G. Caughley, D. Grice e B. Brown. 1983. The distribution and abundance of kangaroos in relation to environment in Western Australia. *Australian Wildlife Research* **10**:435-451.

Smouse, P. E., J. C. Long e R. R. Sokal 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology* **35**:627-632.

Sokal, R. R. e F. J. Rohlf 1995. Biometry, 3rd Edition.W. H. Freeman and Company, New York.

Spitz, F. e S. Leks. 1999. Environmental impact prediction using neural network modelling. An example in wildlife damage. *Journal of Applied Ecology* **36**: 317-326.

Starfield, A. M. e A. L. Bleloch. 1991. Building Models for Conservation and Wildlife Management. Bellwether Press, Edina, Minnesota.

Stern, M. J. 1998. Field comparisons of two rapid vegetation assessment techniques with permanent plot inventory data in Amazonian Peru. Pp. 269-283 In F. Dallmeier e J. A. Comiskey, (eds). Forest Biodiversity Research, Monitoring and Modeling. UNESCO & Parthenon Publishing, Paris.

Stigler, S. M. 1986. The History of Statistics: The Measurement of Uncertainty before 1900. Belknap Press, Cambridge Massachusetts.

Structured-Population Models in Marine, Terrestrial, and Freshwater Systems. Chapman and Hall, New York.

Student. 1908. The probable error of the mean. *Biometrika* **6**:1-25.

ter Braak, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* **86**:1167-1179.

Thornhill, R., A. P. Moller e S. W. Gangestad. 1999. The biological significance of fluctuating asymmetry and sexual selection: a reply to Palmer. *American Naturalist* **154**:234-241.

Tukey, J. W. 1960. Conclusões vs. decisions. *Technometrics* **2**:423-433.

Tukey, J. W. 1972. Some graphic and semigraphic displays. Pp 293-316 In T. A. Bancroft (ed) Statistical Papers in Honour of George W. Snedecor. Iowa State University, Ames.

Tukey, J. W. 1980.We need both exploratory and confirmatory. *American Statistician* **34**:23-25.

Tukey, J. W. 1991. The philosophy of multiple comparisons. *Statistical Science* **6**:100-116.

von Ende, C. N. 1993. Repeated measures analysis: growth and other time-dependent measures. Pp 113-137 In Scheiner, S. M. e J. Gurevitch (eds) Design and Analysis of Ecological Experiments. Chapman and Hall, New York.

Wallace, C. S. e D. L. Dowe. 2000. MML clustering of multistate, Poisson, von Mises circular and Gaussian distributions. *Statistics and Computing* **10**:73-83.

Wardle, G. M. 1998. A graph theory approach to demographic loop analysis. *Ecology* **79**:2539- 2549.

Wartenberg, D., S. Ferson e F. J. Rohlf. 1987. Putting things in order: a critique of detrended correspondence analysis. *American Naturalist* **129**:434-448.

Wilkinson, L. 1999. Statistical methods in psychology journals: guidelines and explanations. *American Psychologist* **54**:594-604.

Williams, B. K. 1983. Some observations on the use of discriminant analysis in ecology. *Ecology* **64**:1283-1291.

Williams, B. K. e K. Titus 1988. Assessment of sampling stability in ecological applications of discriminant analysis. *Ecology* **69**:1275-1291.

Winer, B. J., D. R. Brown e K. M. Michaels. 1991. Statistical Principles in Experimental Design. McGraw-Hill, New York.

Wood, S. N. 1997. Inverse problems and structured-population dynamics. Pp 555-586 In Tuljapurkar, S.e H. Caswell (eds)

Yoccoz, N. G. 1991. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. *Bulletin of the Ecological Society of America* **72**:106-111.

Zar, J. H. 1996. Biostatistical Analysis. 4th ed. Prentice-Hall , London